https://speakerdeck.com/rmoff/

What is Apache Kafka, and What is a Streaming Platform?

Budapest Data Forum, 14 Jun 2018

Robin Moffatt

@rmoff robin@confluent.io

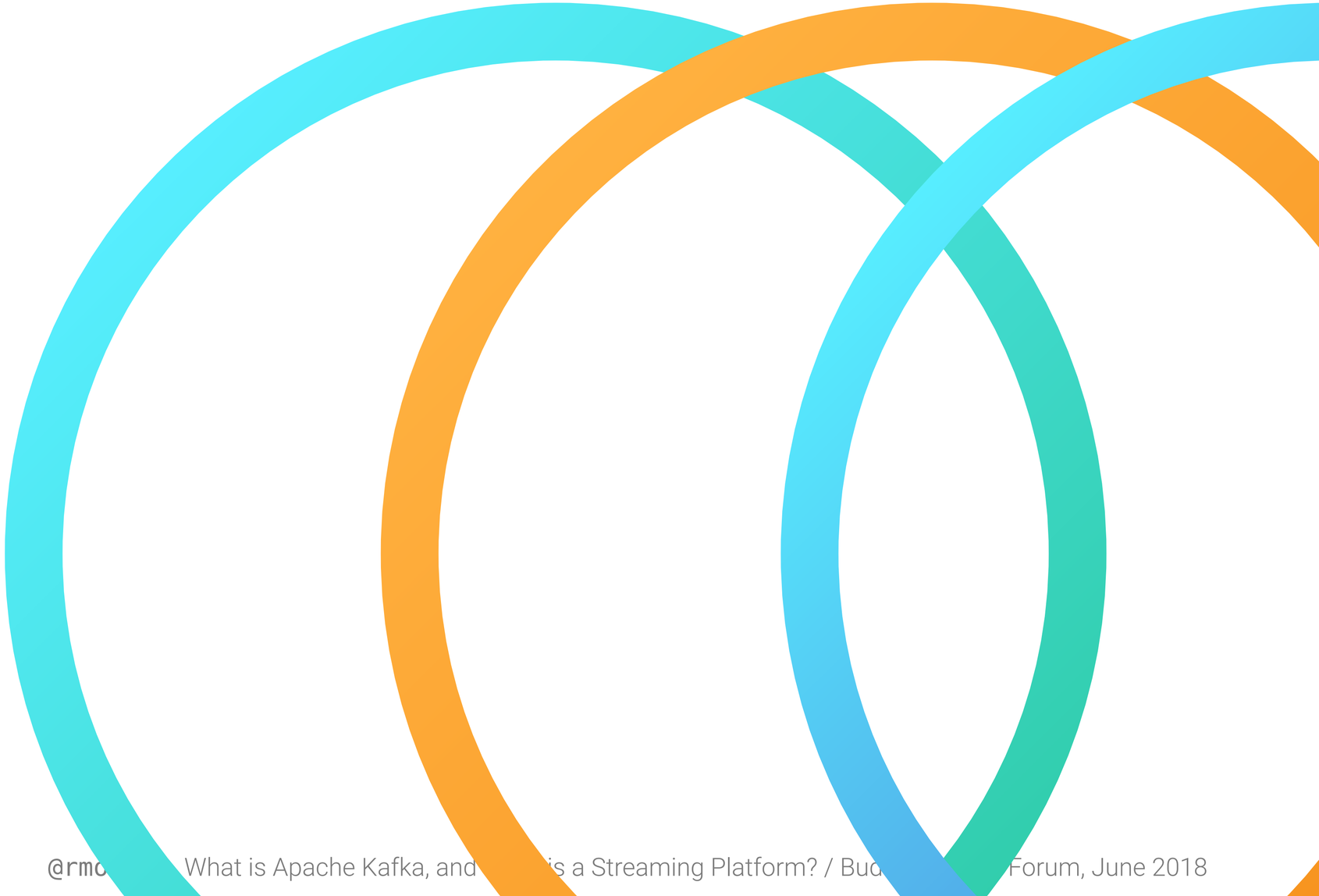
$ whoami

- Developer Advocate @ Confluent
- Working in data & analytics since 2001
- Oracle ACE Director & Dev Champion
- Blogging : http://rmoff.net & http://cnfl.io/rmoff
- Twitter: **@rmoff**
  - Geek stuff
  - Beer & Fried Breakfasts

https://speakerdeck.com/rmoff/

“Apache Kafka is a Streaming Platform

Three Lenses



What is Apache Kafka?

01

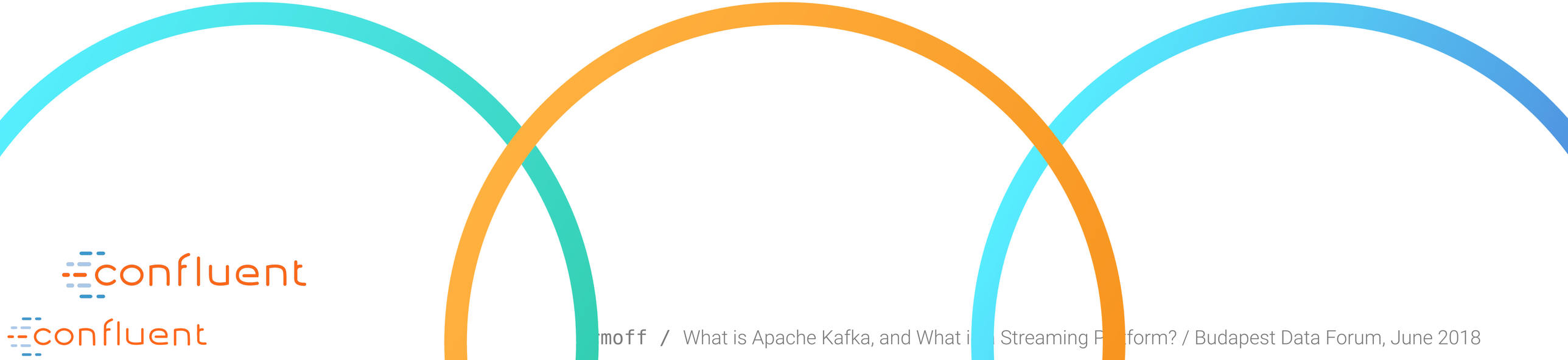
Messaging
Done Right

02

Scalable Streaming
Data Pipelines

03

Foundation for
Stream Processing



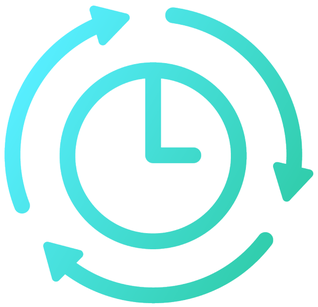
confluent

confluent

Lens 1: Messaging Done Right

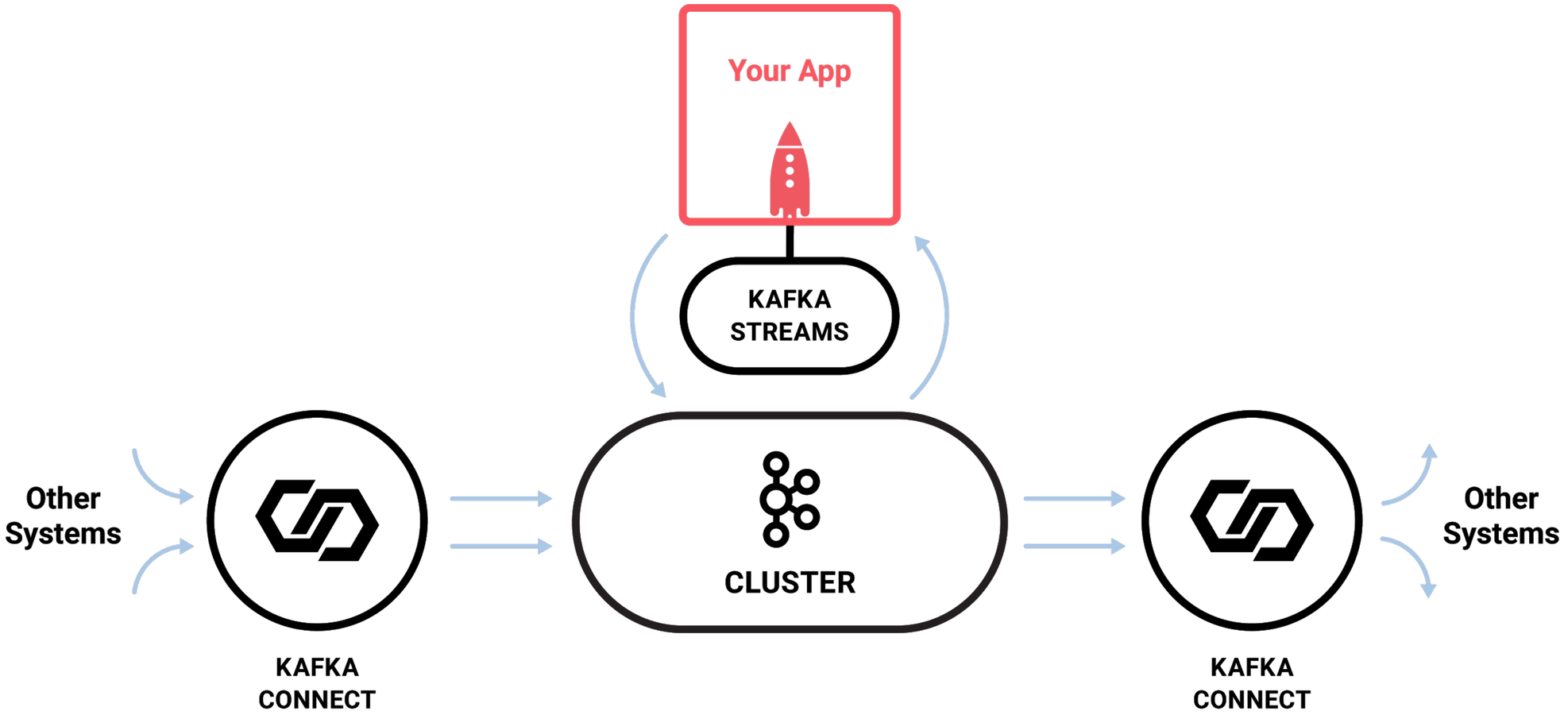
Scalability

True Storage

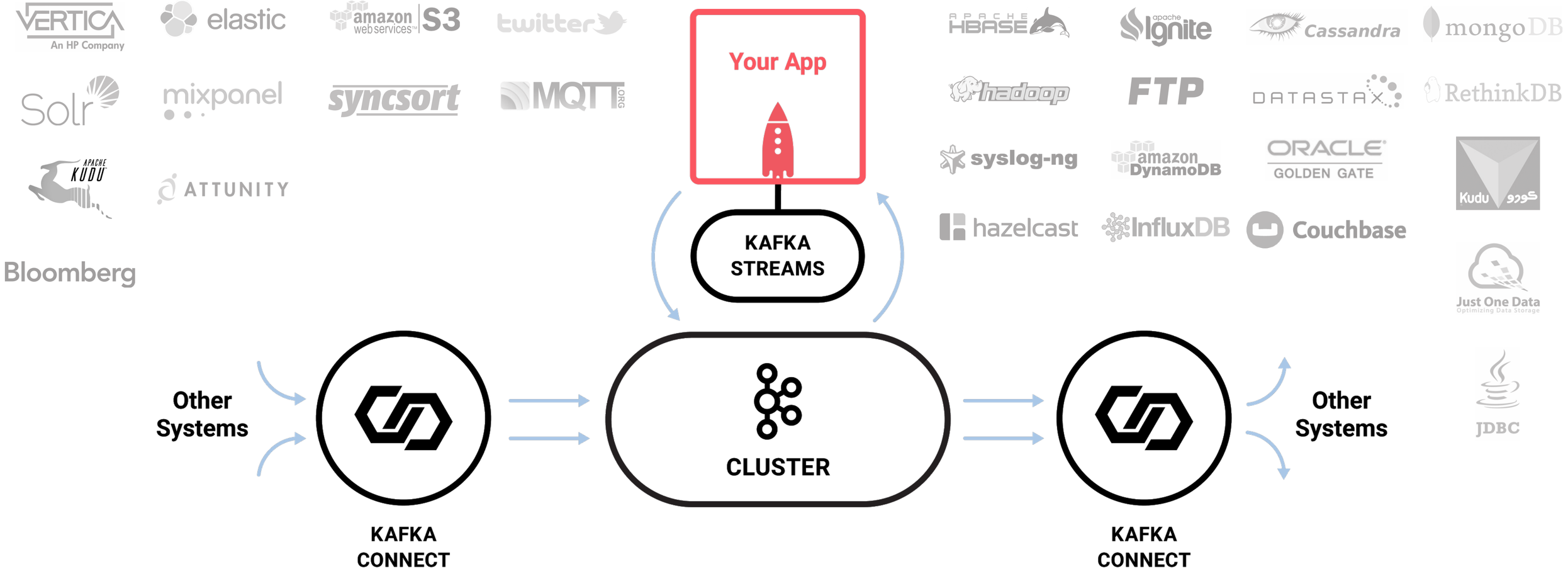


Real-Time Processing

Lens 2: Scalable Streaming Data Pipelines

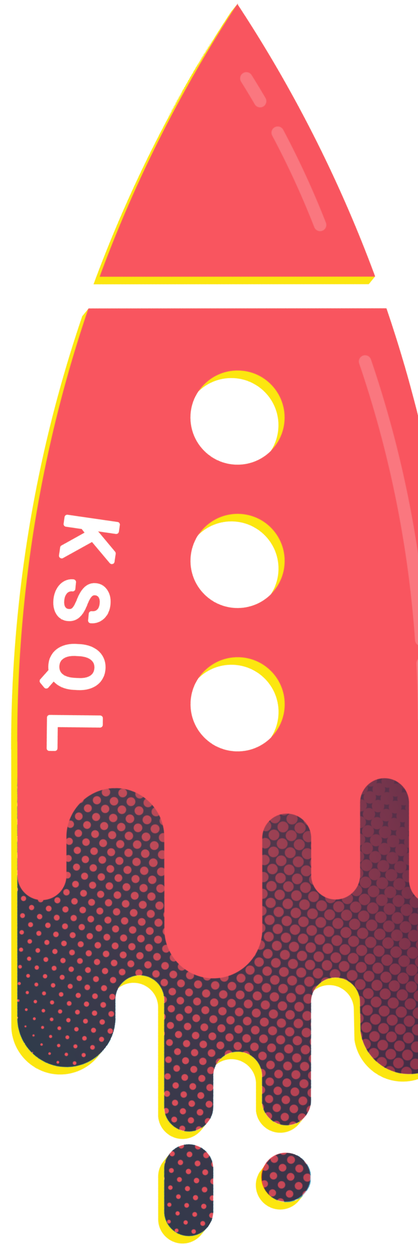


Lens 2: Scalable Streaming Data Pipelines

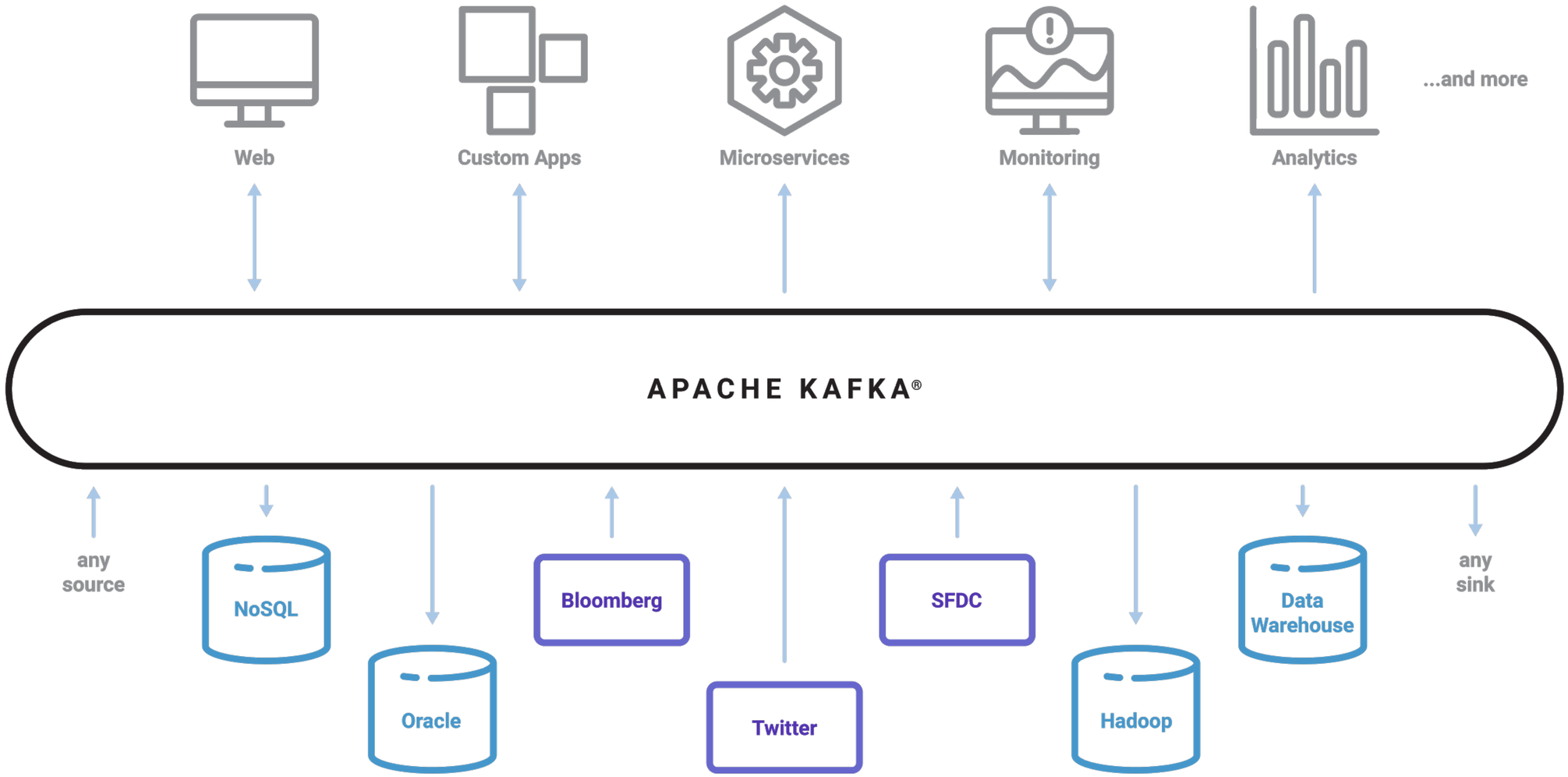


Lens 3: Foundation for Stream Processing

KSQL
is the
Streaming
SQL Engine
for
Apache Kafka



The Streaming Platform

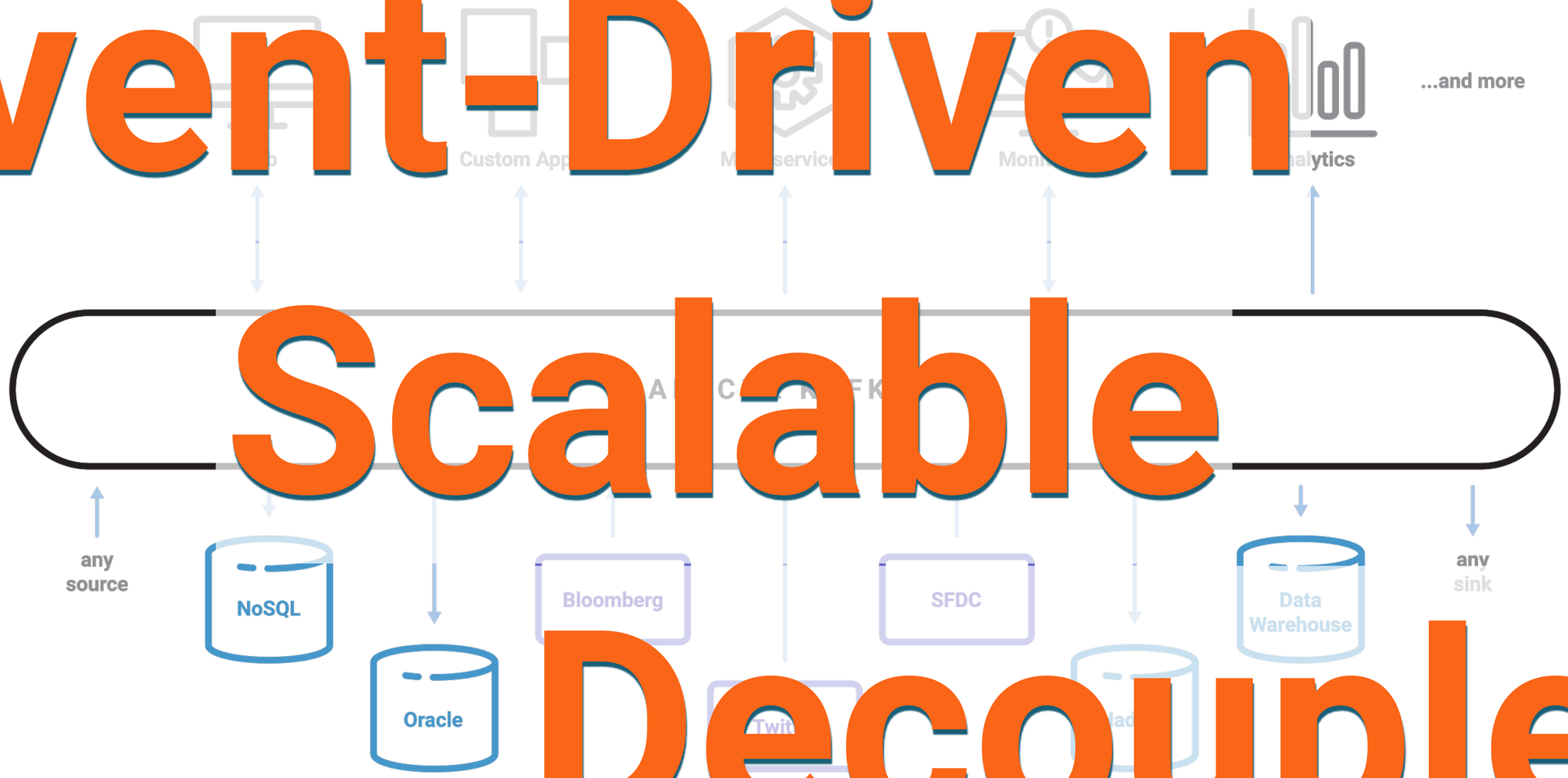


The Streaming Platform

Event-Driven

Scalable

Decoupled



"Bold claim: all your data is event streams

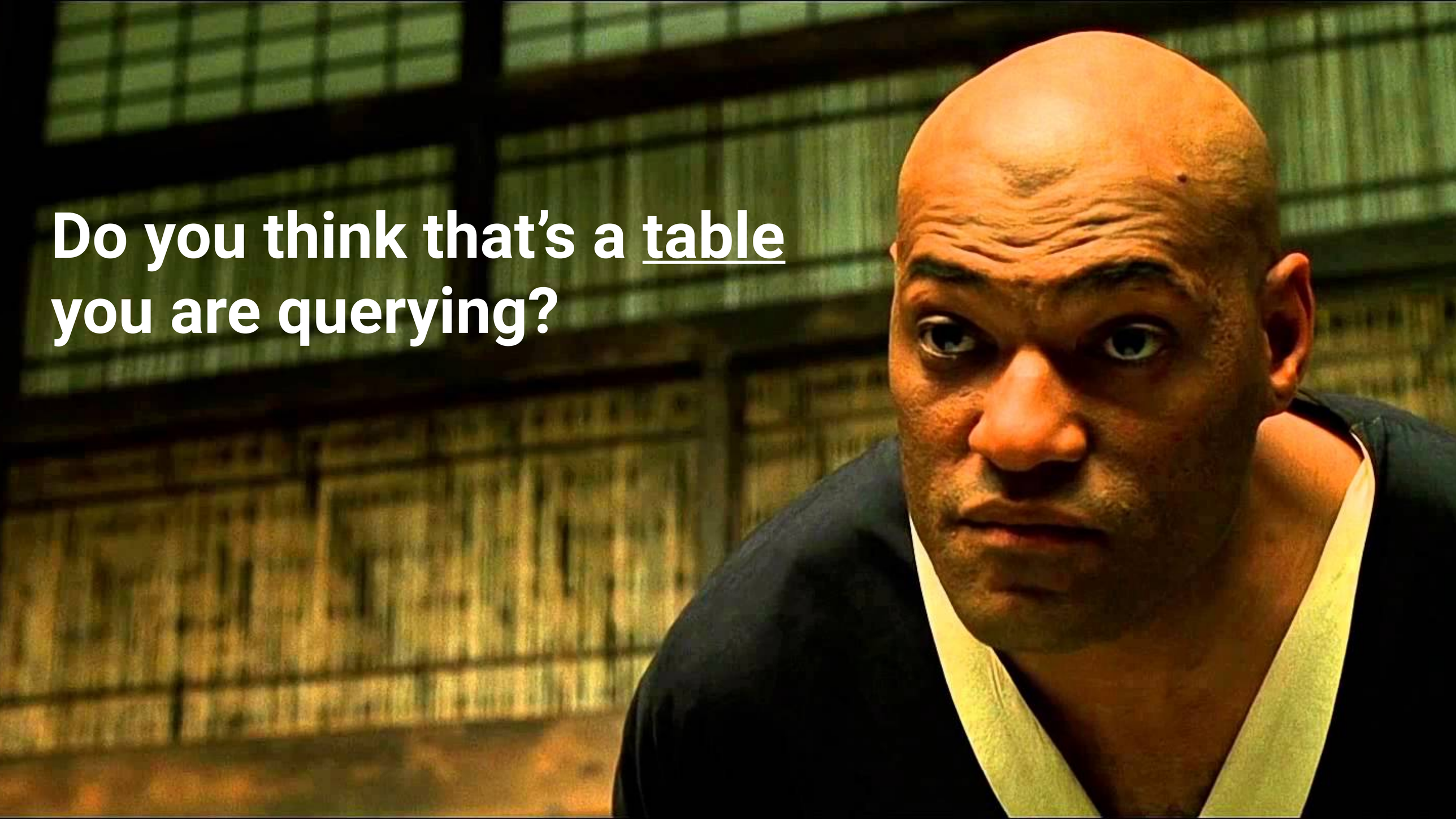
A Customer Experience

A Sale

A Sensor Reading

An Application Log Entry

Databases

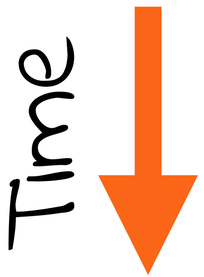

Do you think that’s a table
you are querying?

The Stream-Table Duality

Account ID	Balance
12345	€50

The Stream-Table Duality

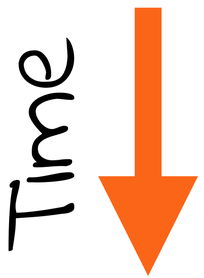
Account ID	Amount
12345	+ €50



Account ID	Balance
12345	€50

The Stream-Table Duality

Account ID	Amount
12345	+ €50
12345	+ €25



Account ID	Balance
12345	€50

Account ID	Balance
12345	€75

The Stream-Table Duality

Time

Account ID	Amount
12345	+ €50
12345	+ €25
12345	-€60

Account ID	Balance
12345	€50

Account ID	Balance
12345	€75

Account ID	Balance
12345	€15

The Stream-Table Duality

Stream

Time

Account ID	Amount
12345	+ €50
12345	+ €25
12345	-€60

Table

Account ID	Balance
12345	€50

Account ID	Balance
12345	€75

Account ID	Balance
12345	€15

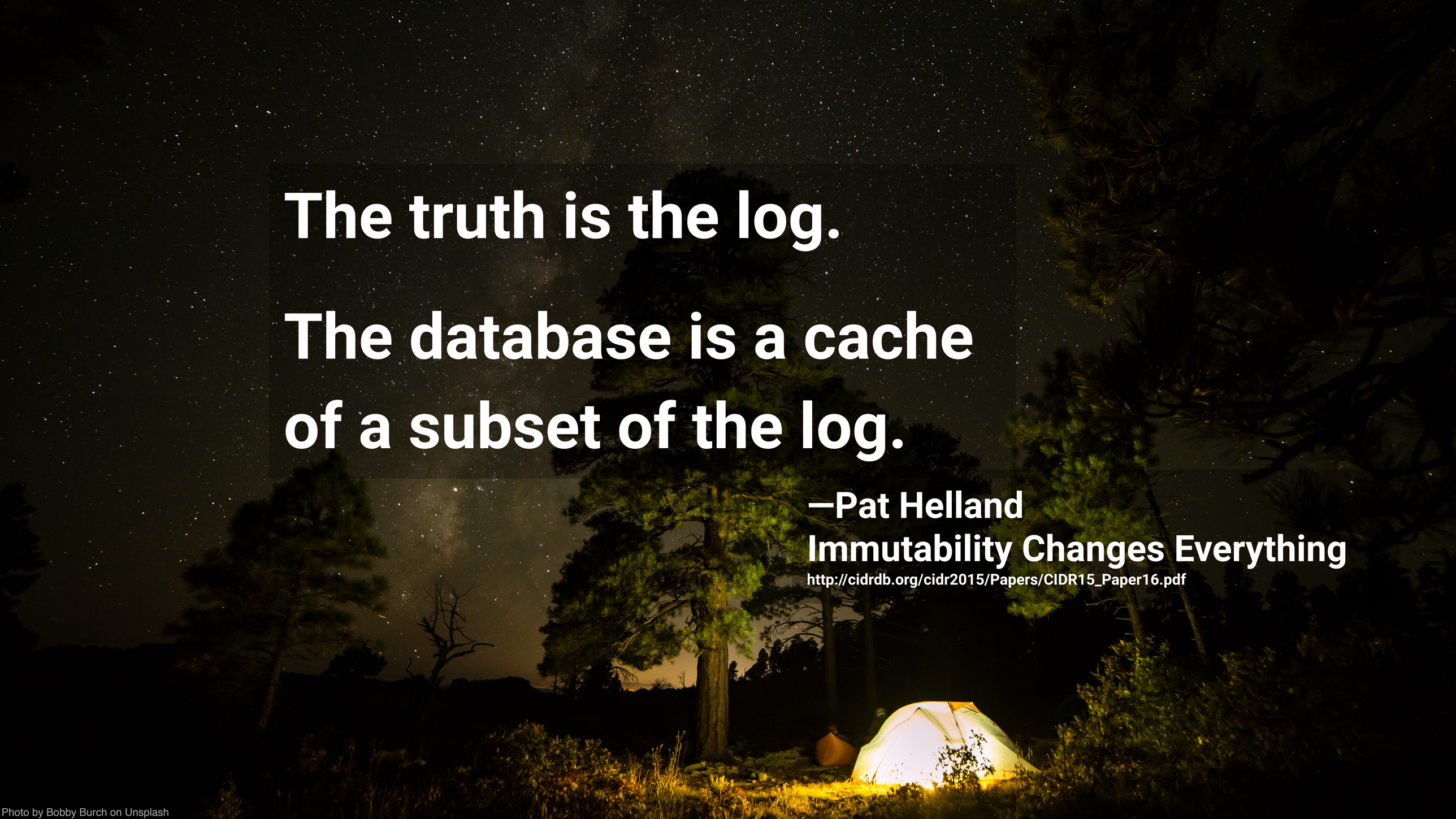
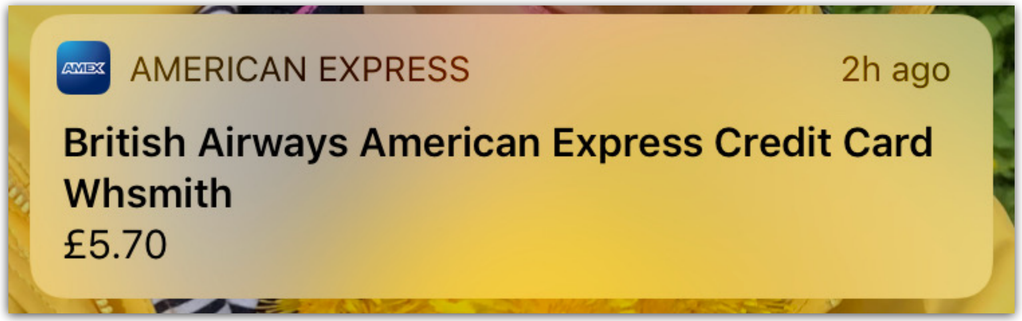


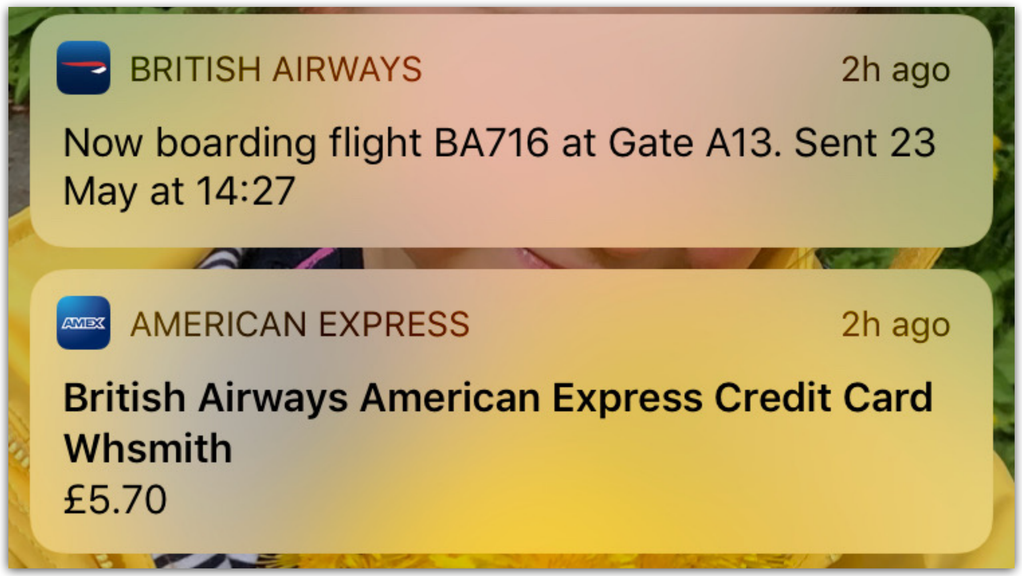
Photo by Bobby Burch on Unsplash

Event-Driven architectures in action...

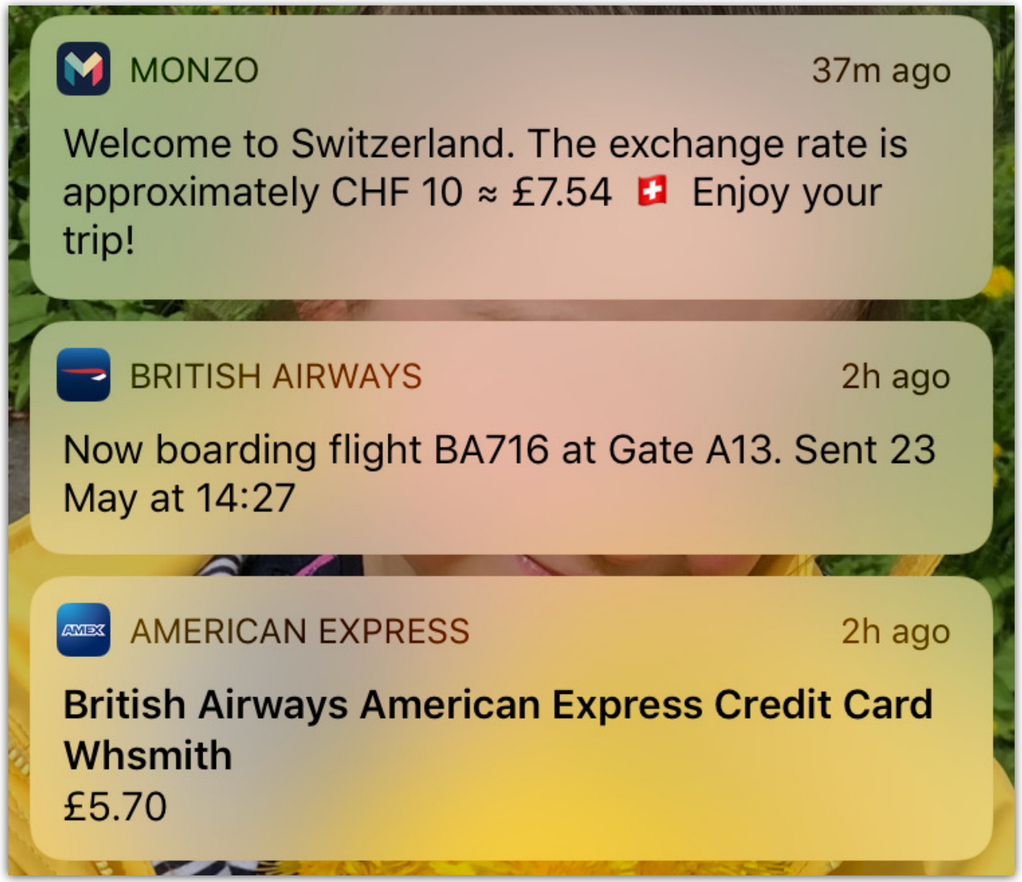
Event-Driven architectures in action...



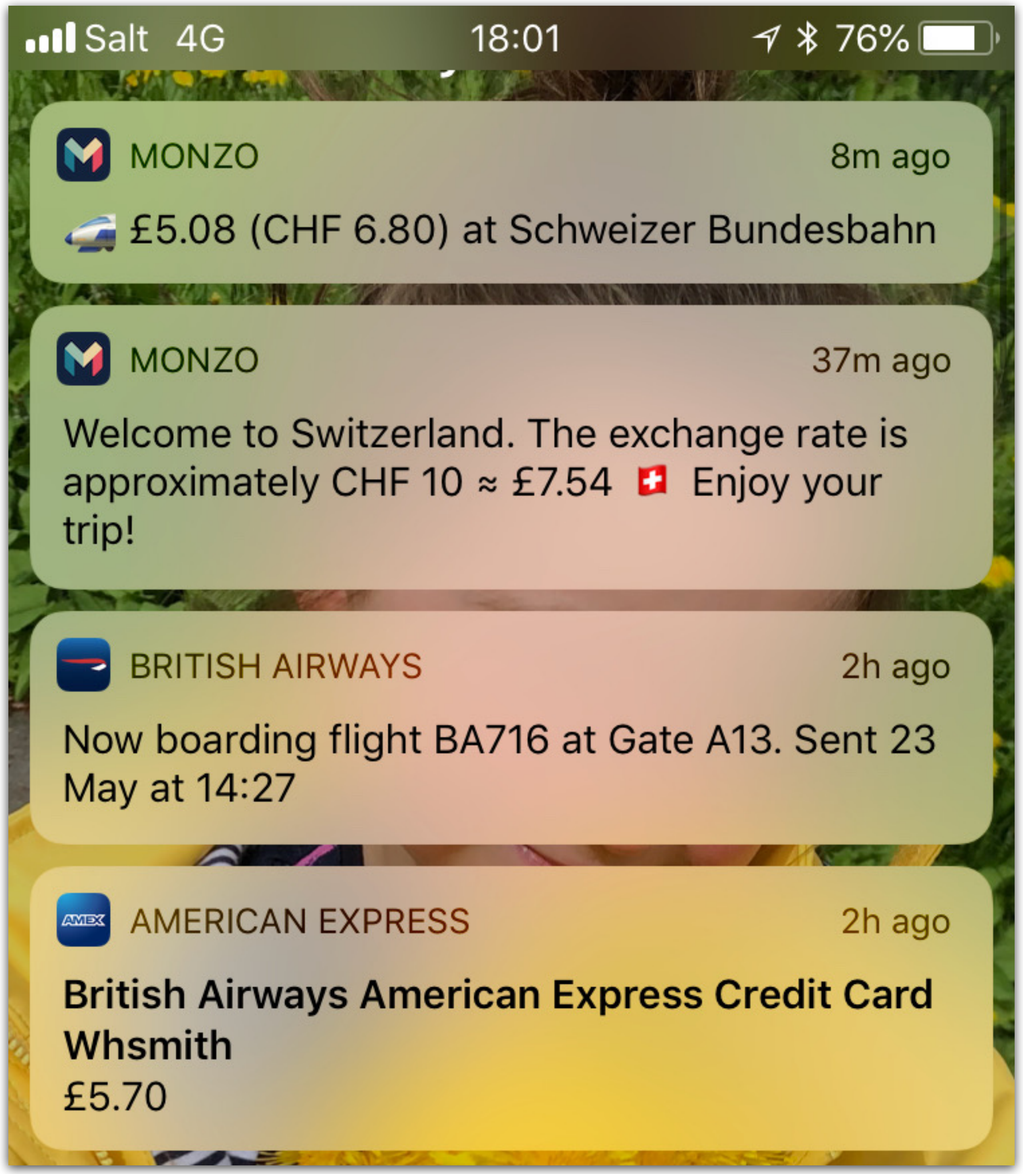
Event-Driven architectures in action...



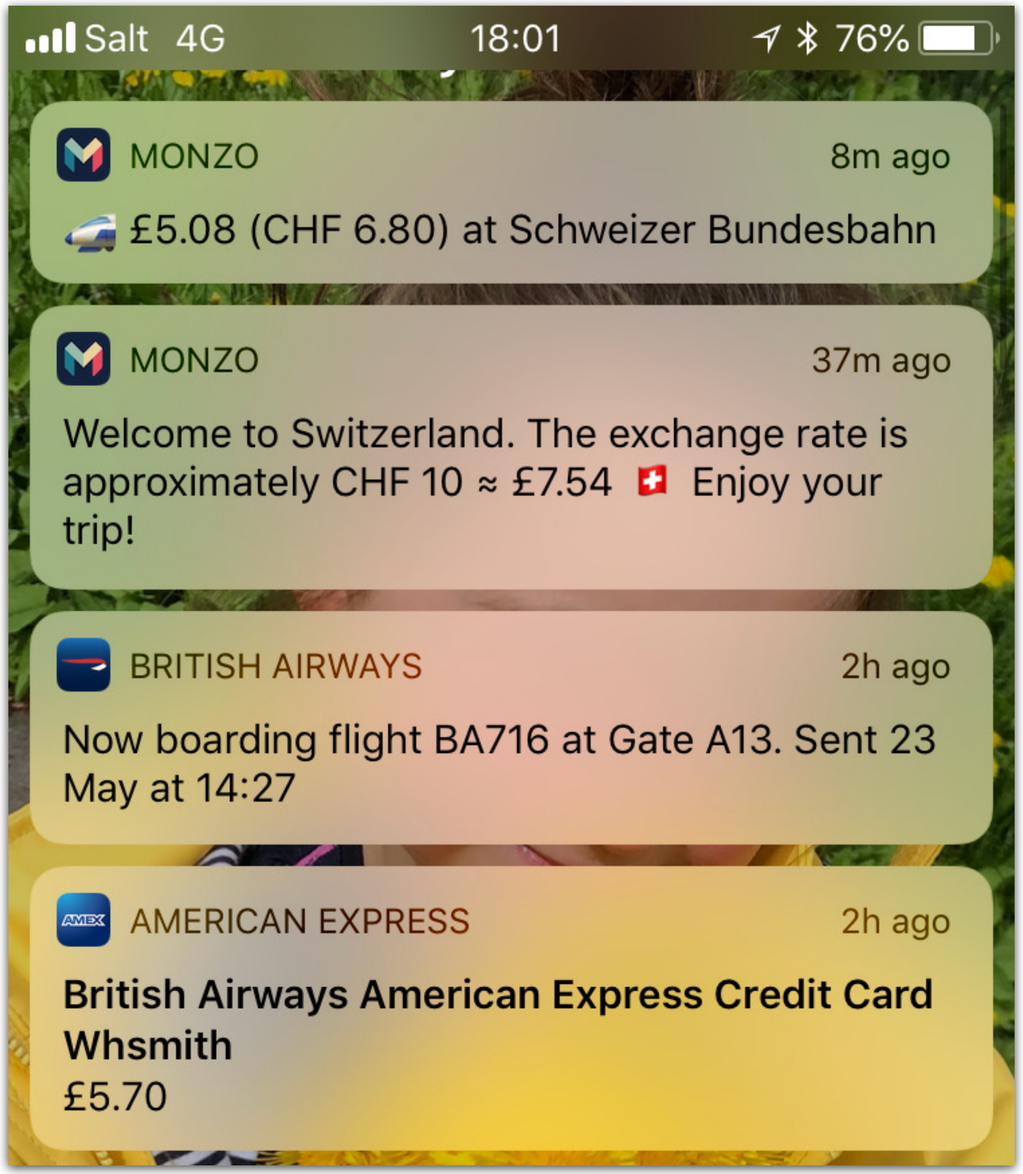
Event-Driven architectures in action…

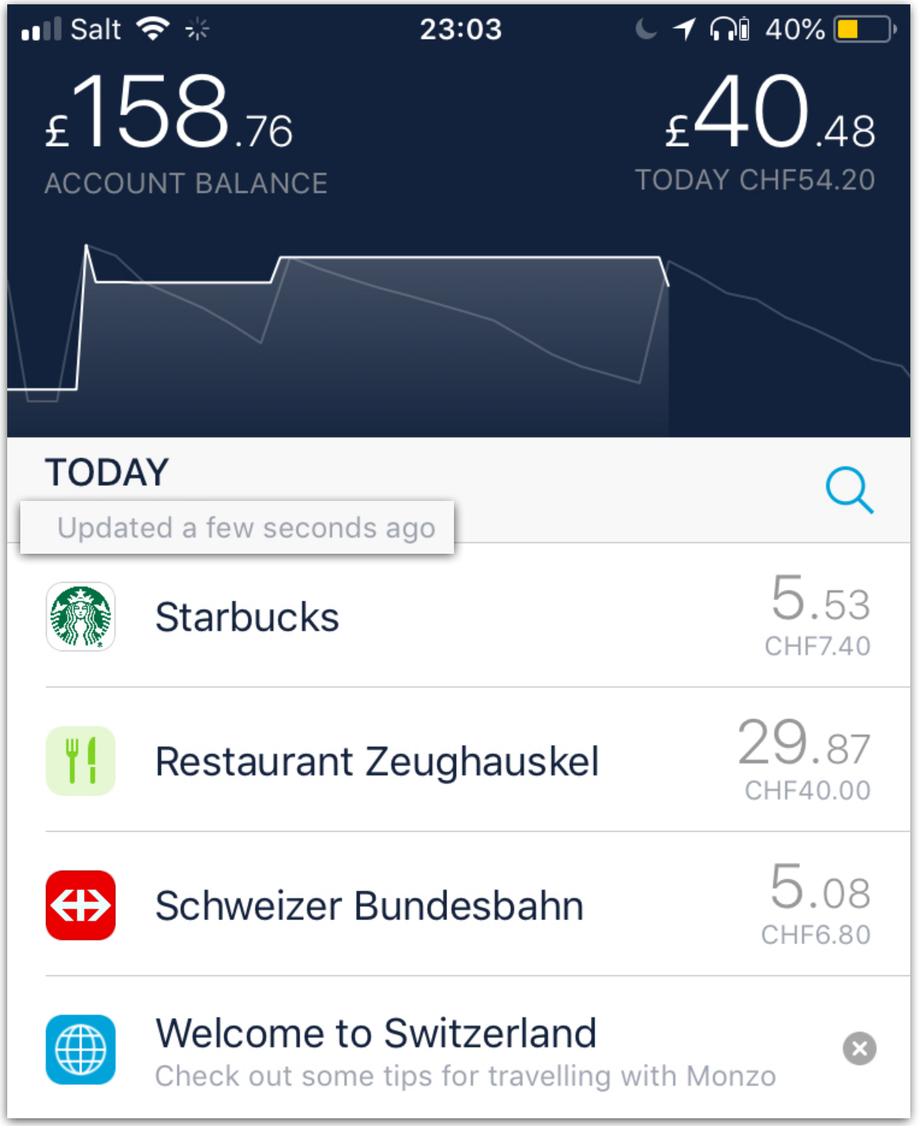


Event-Driven architectures in action…

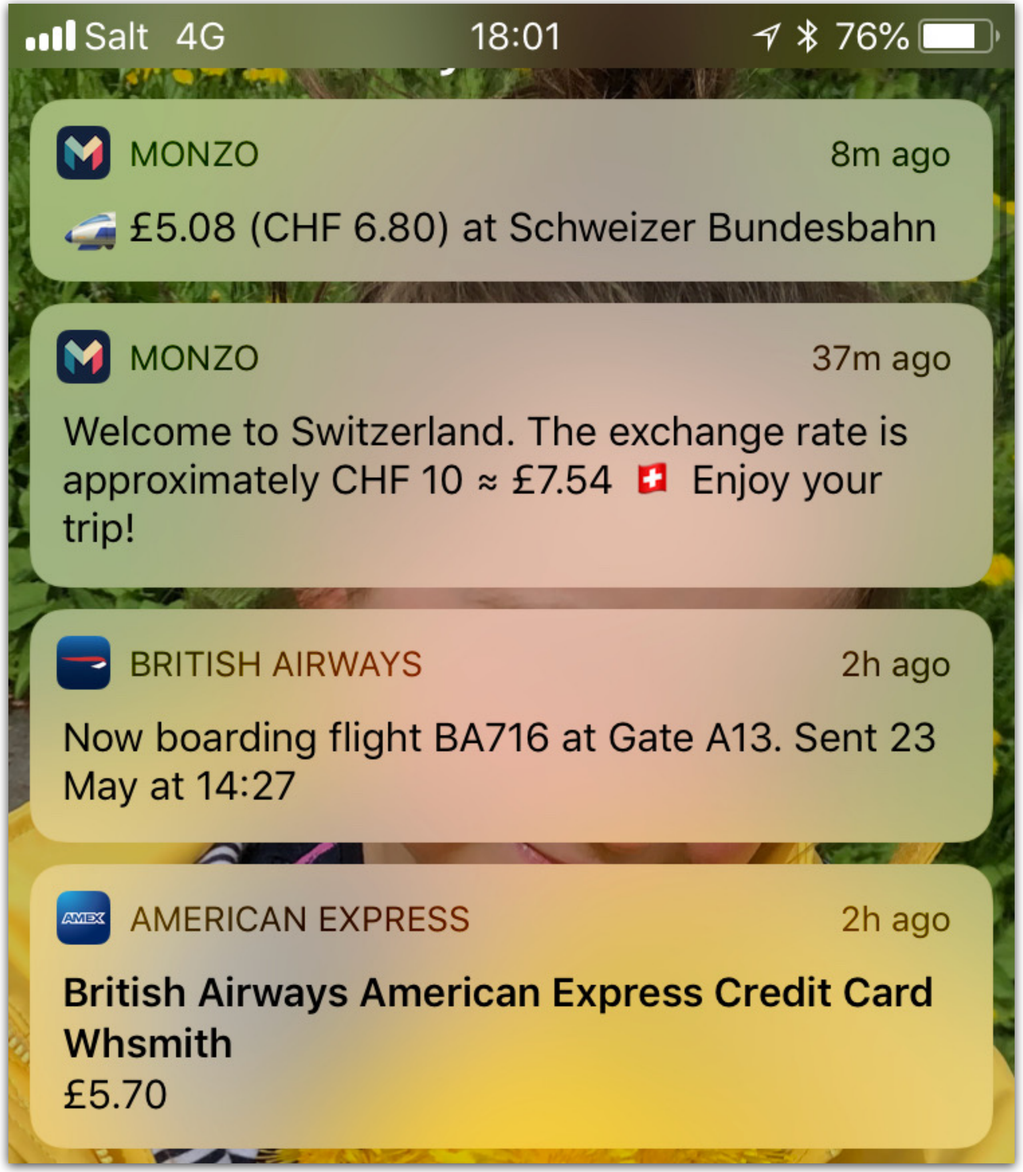


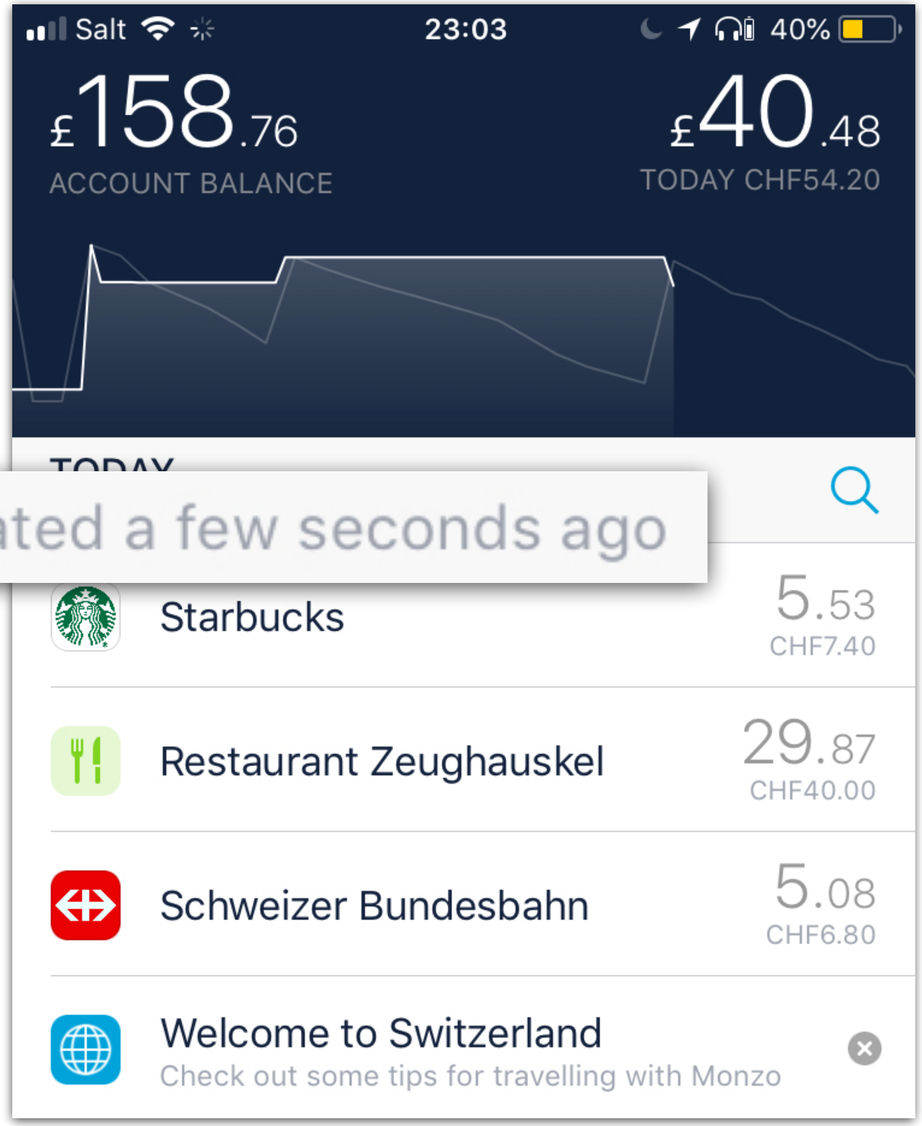
Event-Driven architectures in action…





Event-Driven architectures in action...

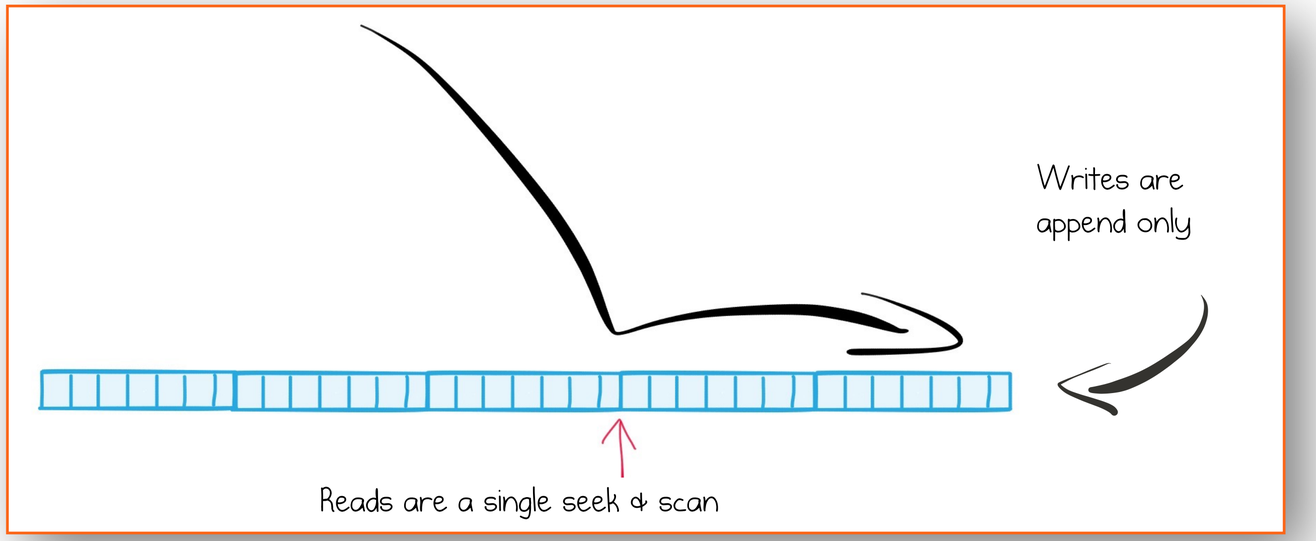




A Brief Look at Kafka's Technology

Kafka

A **Distributed Commit Log**. Publish and subscribe to streams of records. Highly scalable, high throughput. Supports transactions. Persisted data. Stream processing.



Producer & Consumer APIs

Open-source client libraries for numerous languages, to directly integrate with your applications.

python™

C++

Go

Microsoft® .NET™

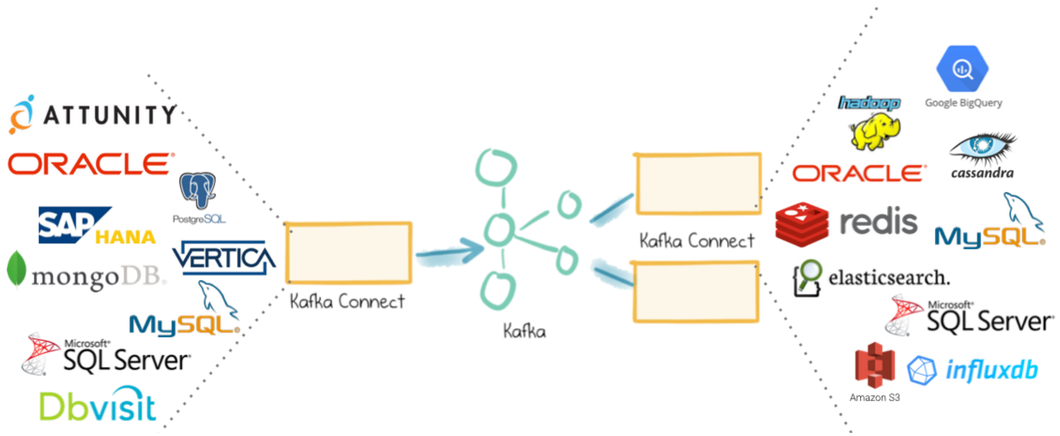
Scala

C

Java™

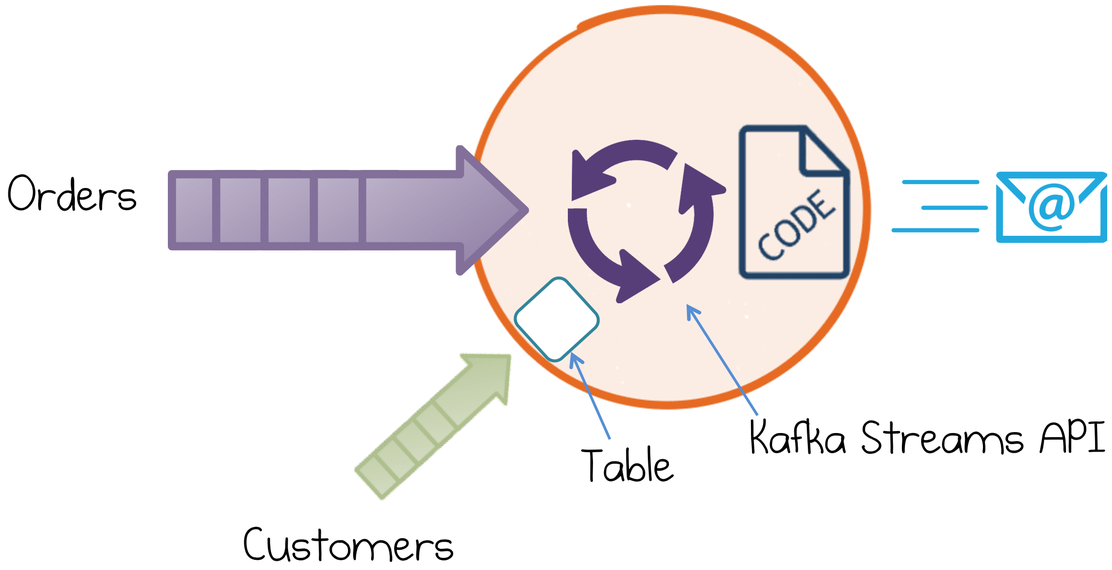
Kafka Connect API

Reliable and scalable integration of Kafka with other systems – no coding required.

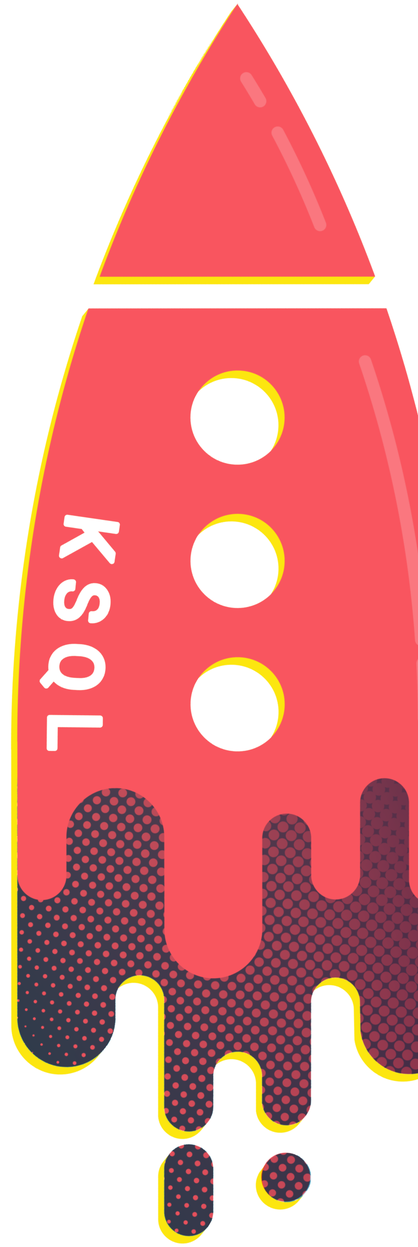


Kafka Streams API

Write standard Java applications & microservices to process your data in real-time



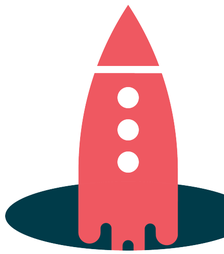
KSQL
is the
Streaming
SQL Engine
for
Apache Kafka



KSQL for Real-Time Monitoring

- Log data monitoring, tracking and alerting
- syslog data
- Sensor / IoT data

```
CREATE STREAM SYSLOG_INVALID_USERS AS
SELECT HOST, MESSAGE
FROM SYSLOG
WHERE MESSAGE LIKE '%Invalid user%';
```

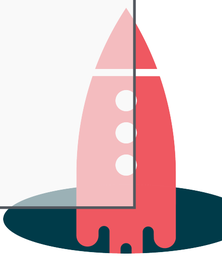


http://cnfl.io/syslogs-filtering / http://cnfl.io/syslog-alerting

KSQL for Streaming ETL

Joining, filtering, and aggregating streams of event data

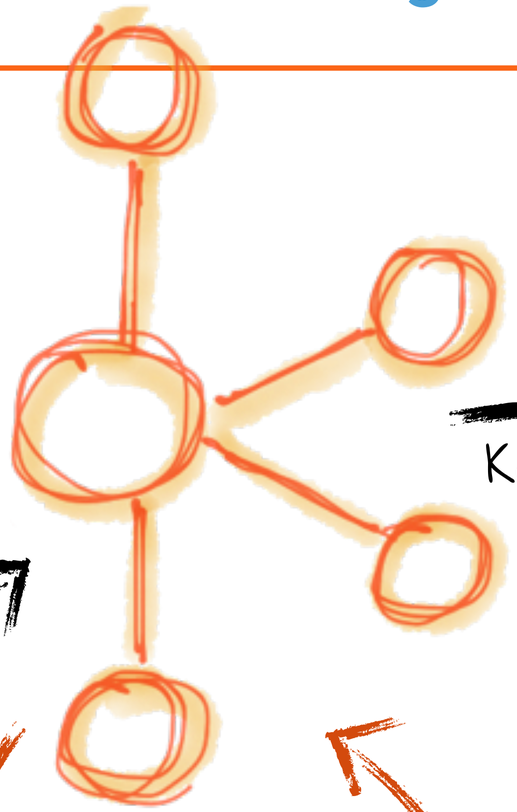
```
CREATE STREAM platinum_customer_ratings AS
 SELECT r.message, r.rating,
        c.customer_name, c.level
  FROM ratings r
  LEFT JOIN customers c
    ON r.userid = c.id
  WHERE c.level = 'Platinum';
```



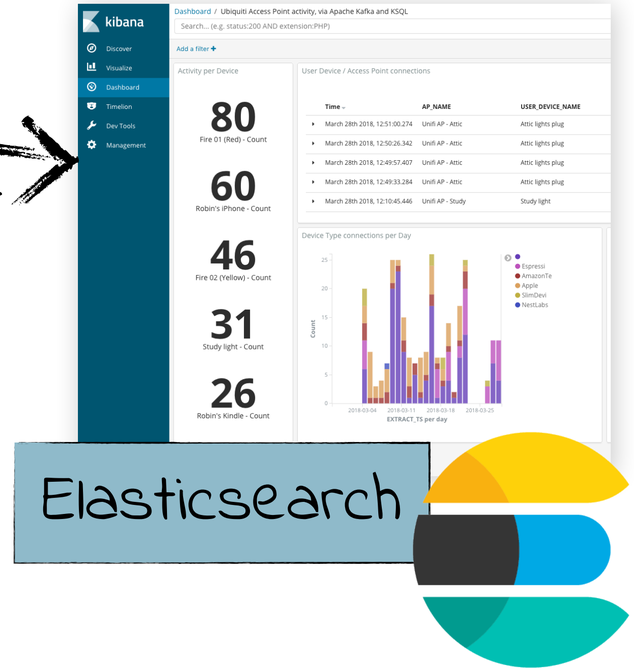
Streaming ETL with Apache Kafka and KSQL


```
{
  "rating_id": 2133,
  "user_id": 9,
  "stars": 1,
  "route_id": 7219,
  "rating_time": 1519402815063,
  "channel": "web",
  "message": "worst. flight. ever. #neveragain"
}
```


Producer API



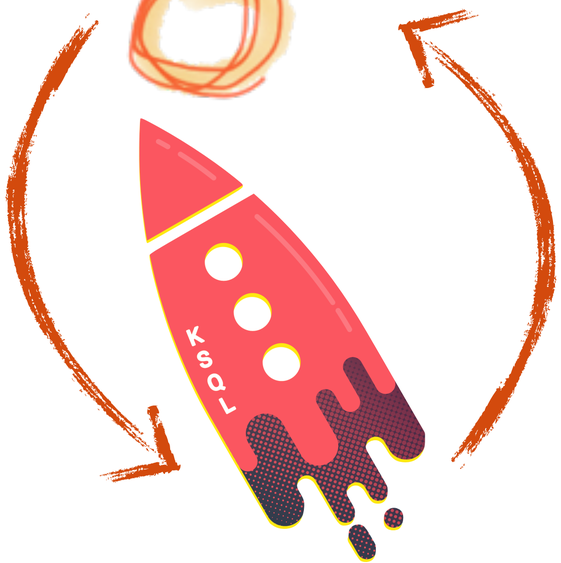
Kafka Connect



Elasticsearch

Kafka Connect
Debezium

MySQL



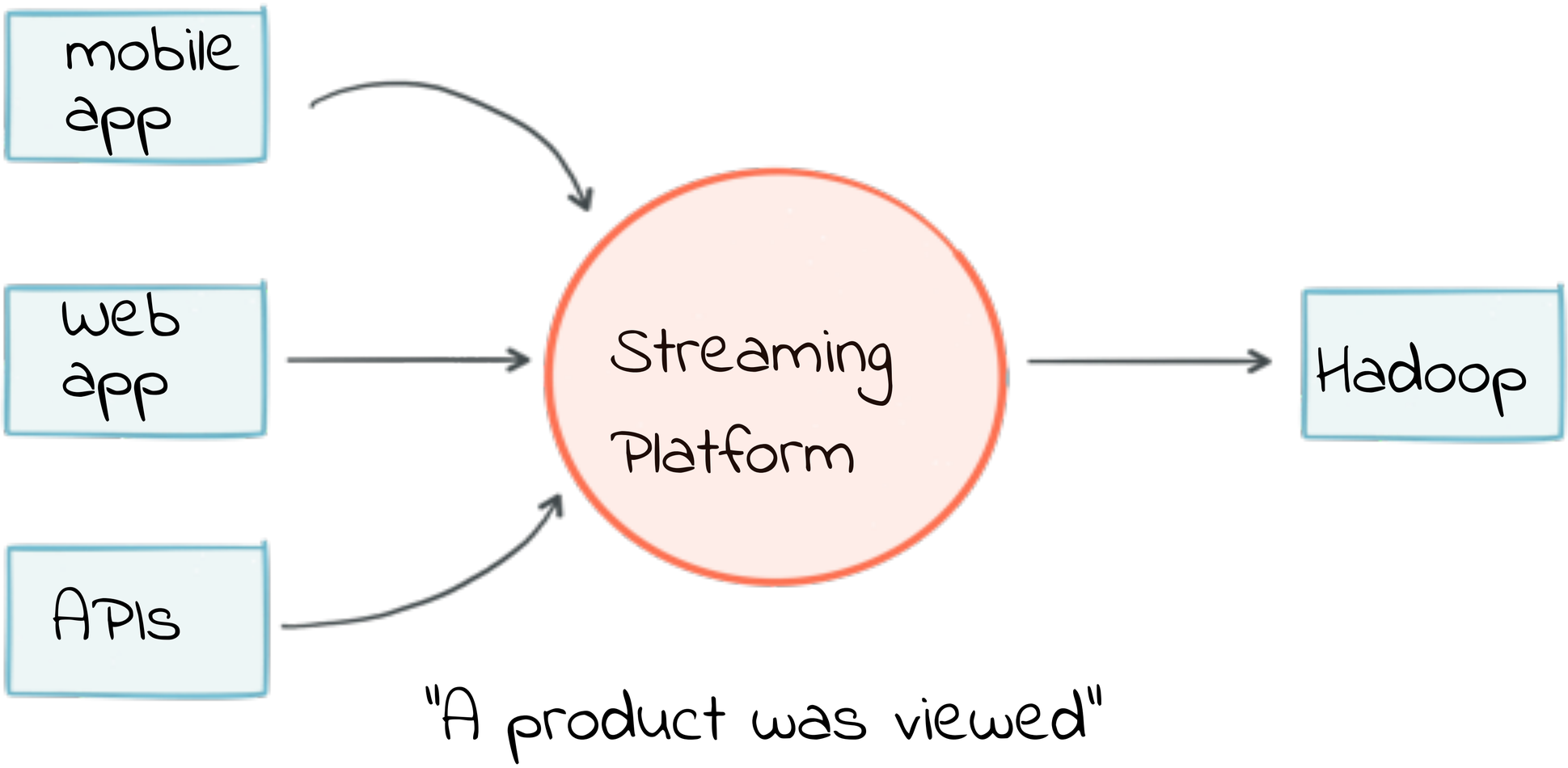

```
{
  "uid": 9,
  "name": "Neha",
  "locale": "en_US",
  "address_city": "Palo Alto",
  "elite": "P"
}
```

“What Problems does Kafka Solve?

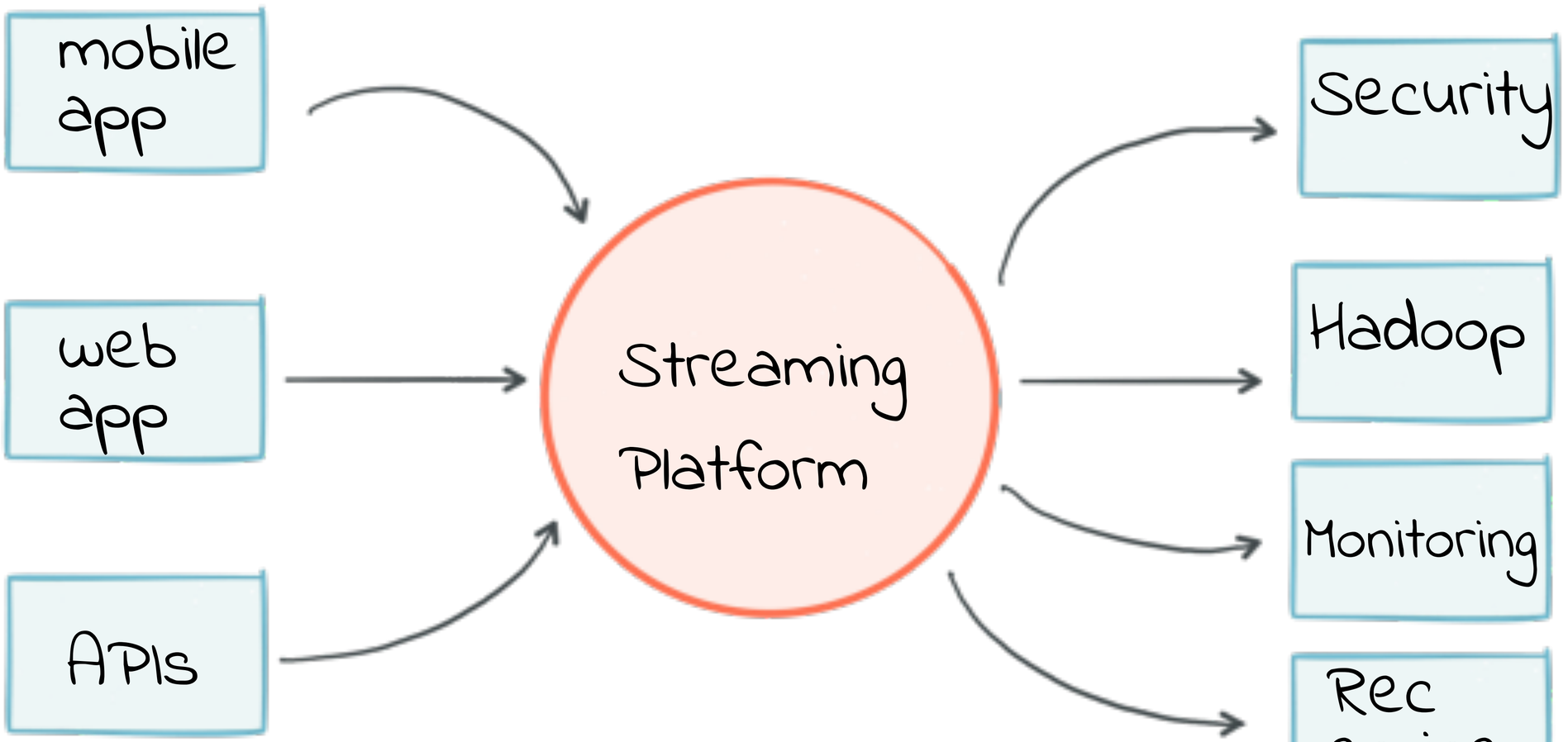
Event-Centric Thinking

"A product was viewed"

Event-Centric Thinking



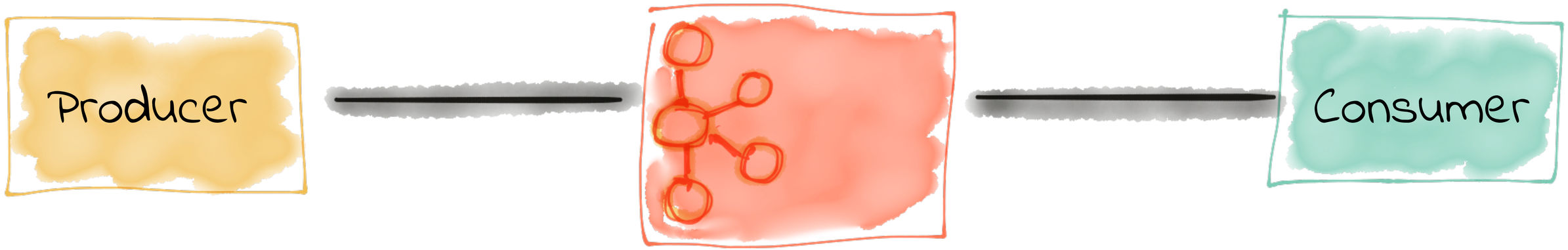
Event-Centric Thinking



"A product was viewed"

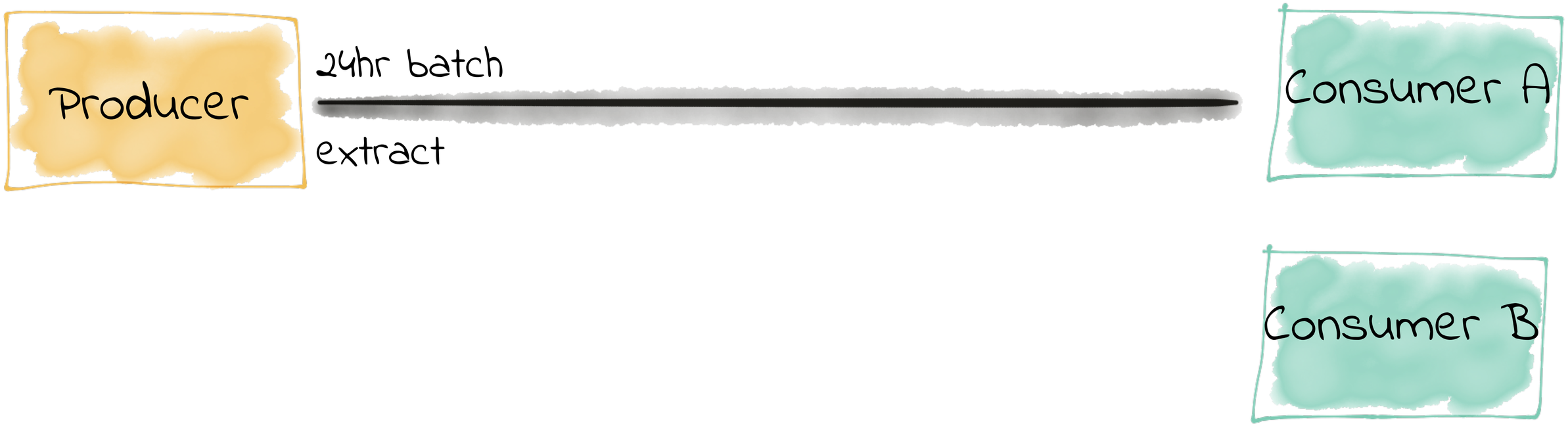
System Availability and Event Buffering

System Availability and Event Buffering

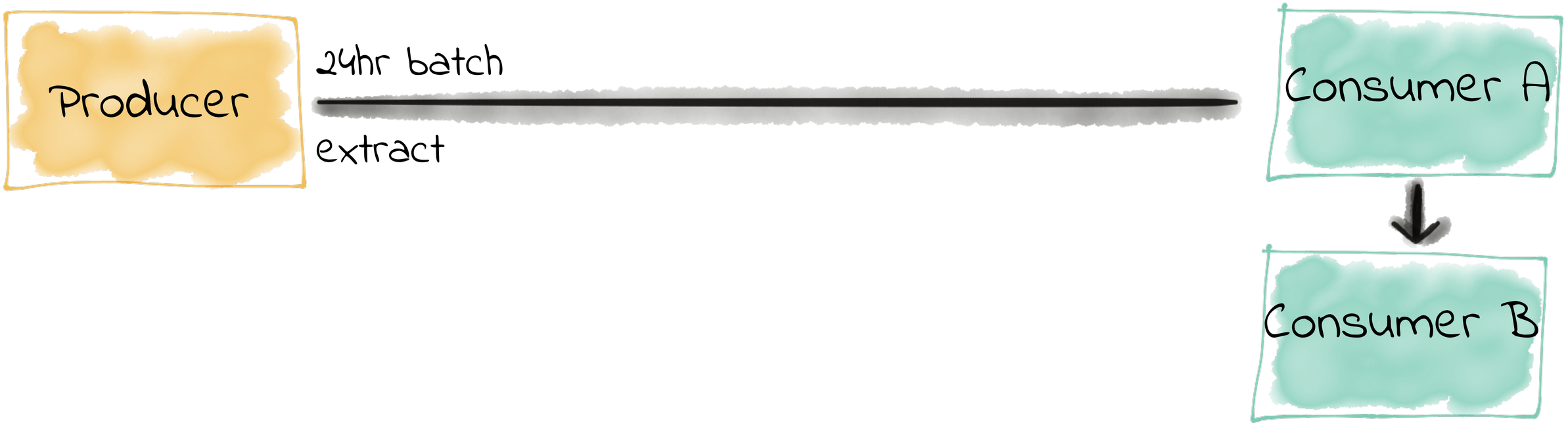


Varying Latency Requirements / Batch vs Stream

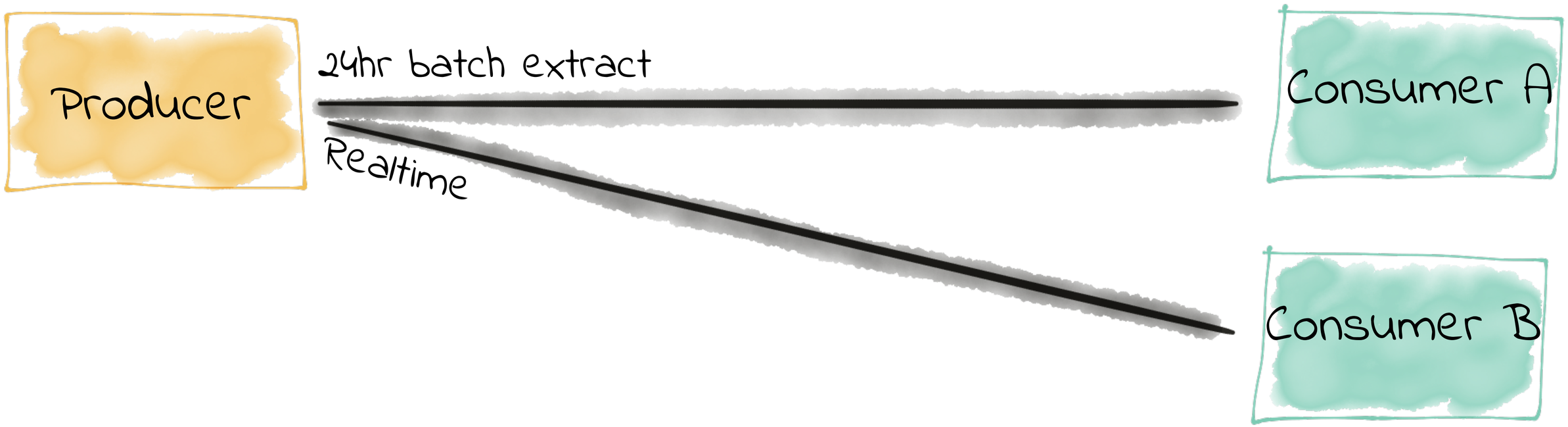
Varying Latency Requirements / Batch vs Stream



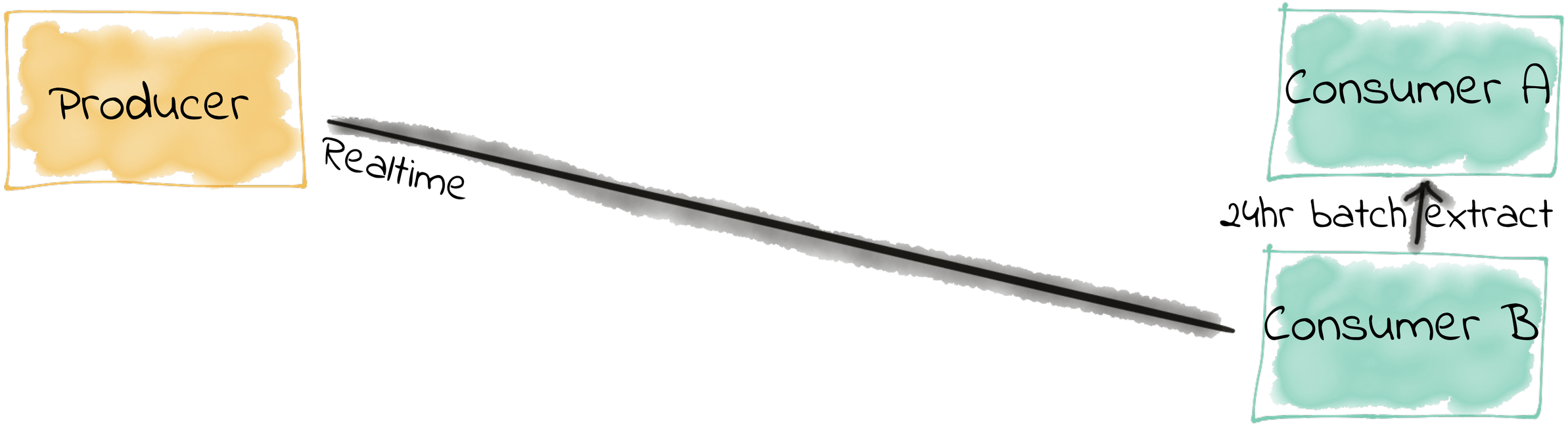
Varying Latency Requirements / Batch vs Stream



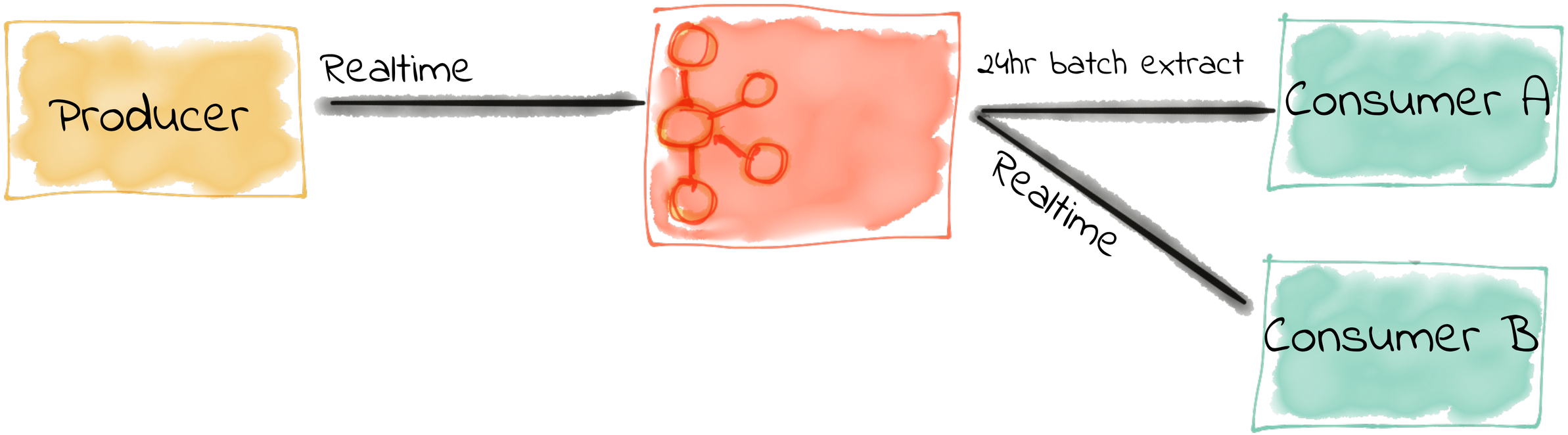
Varying Latency Requirements / Batch vs Stream



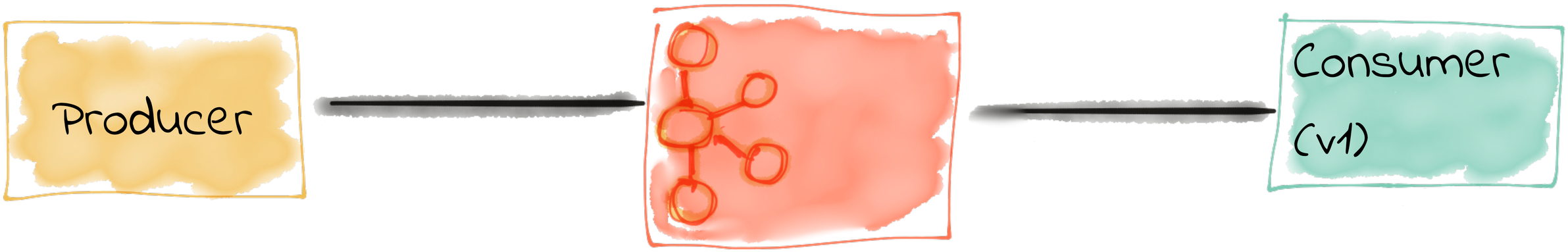
Varying Latency Requirements / Batch vs Stream



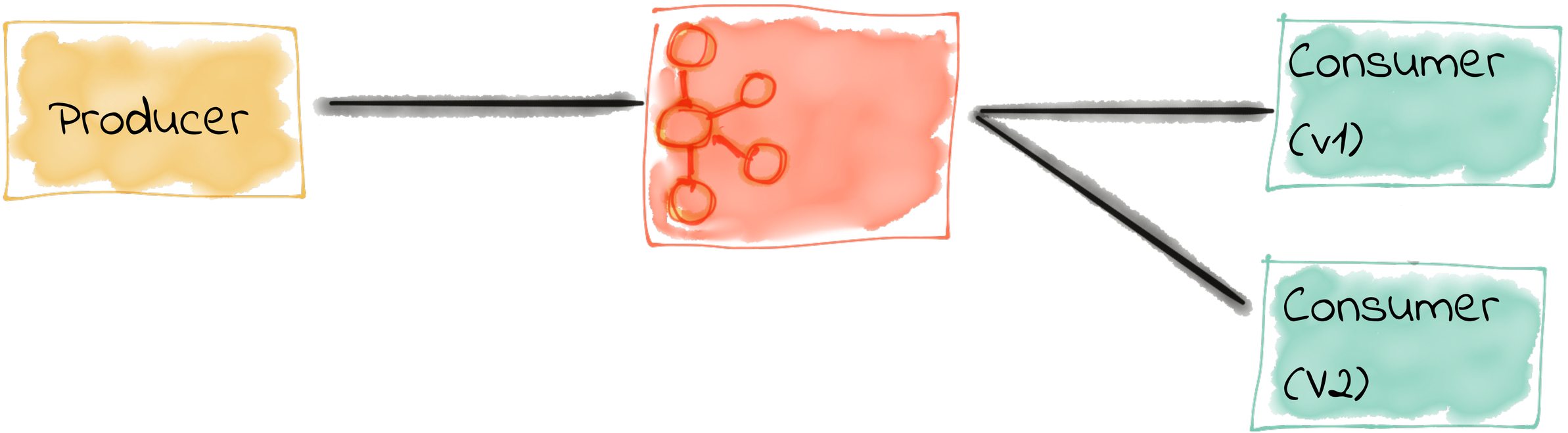
Varying Latency Requirements / Batch vs Stream



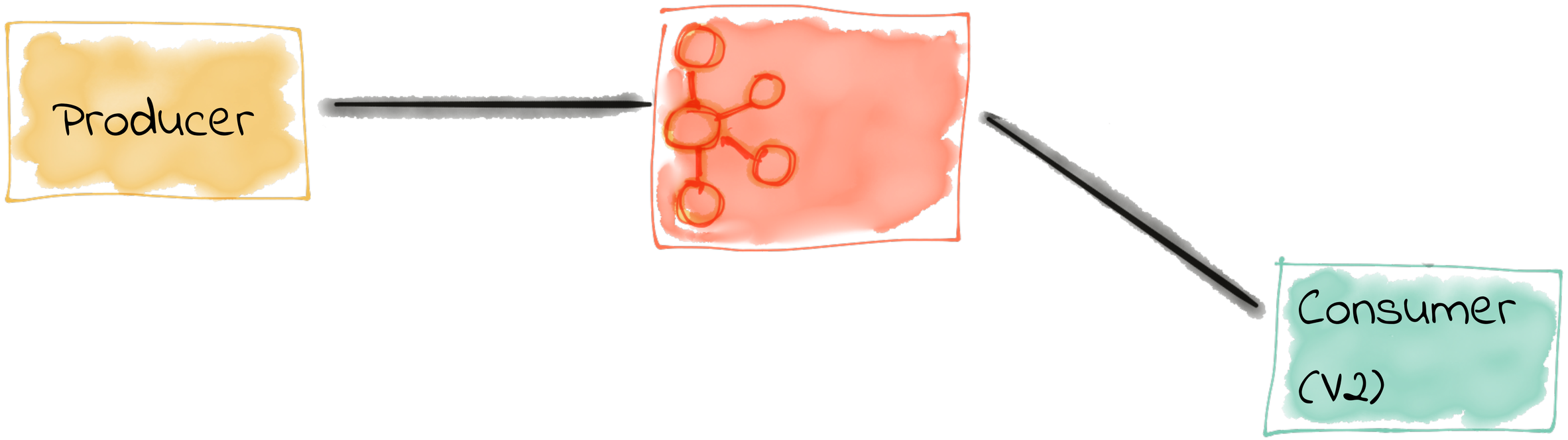
Technology & Code/Algo version Changes



Technology & Code/Algo version Changes



Technology & Code/Algo version Changes

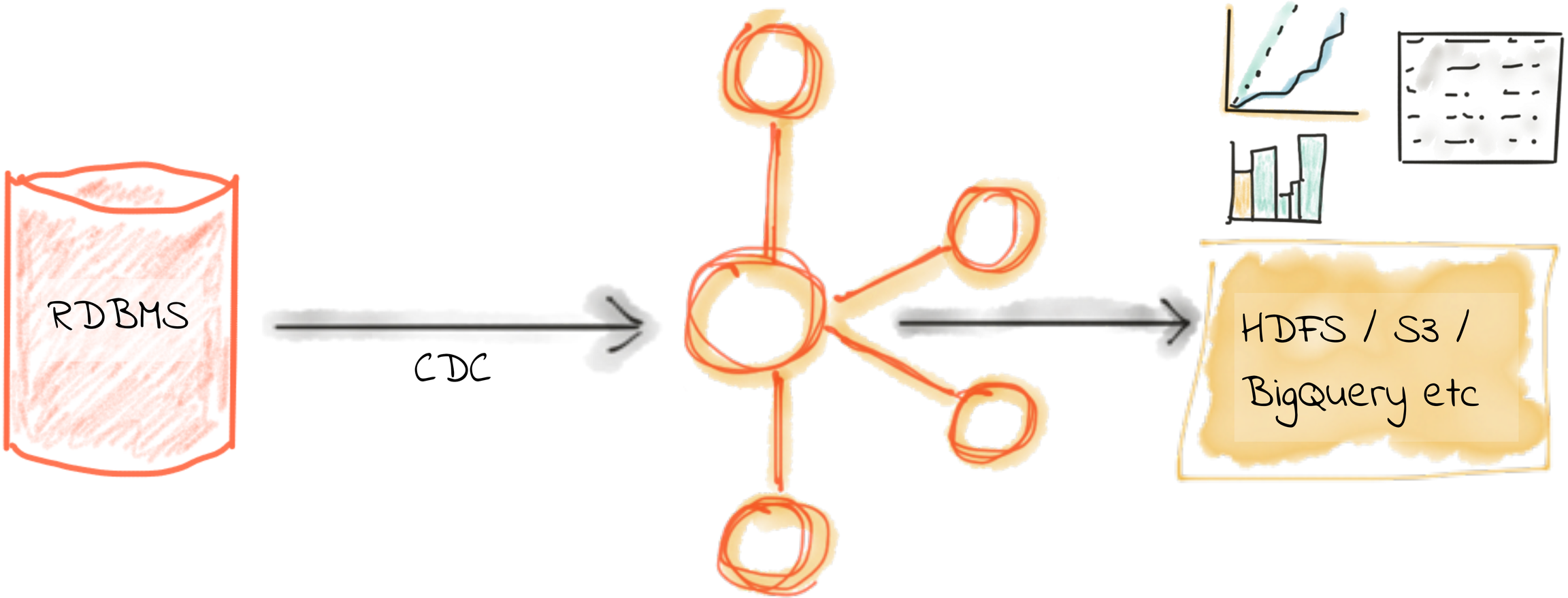


“Architectural Patterns with Apache Kafka

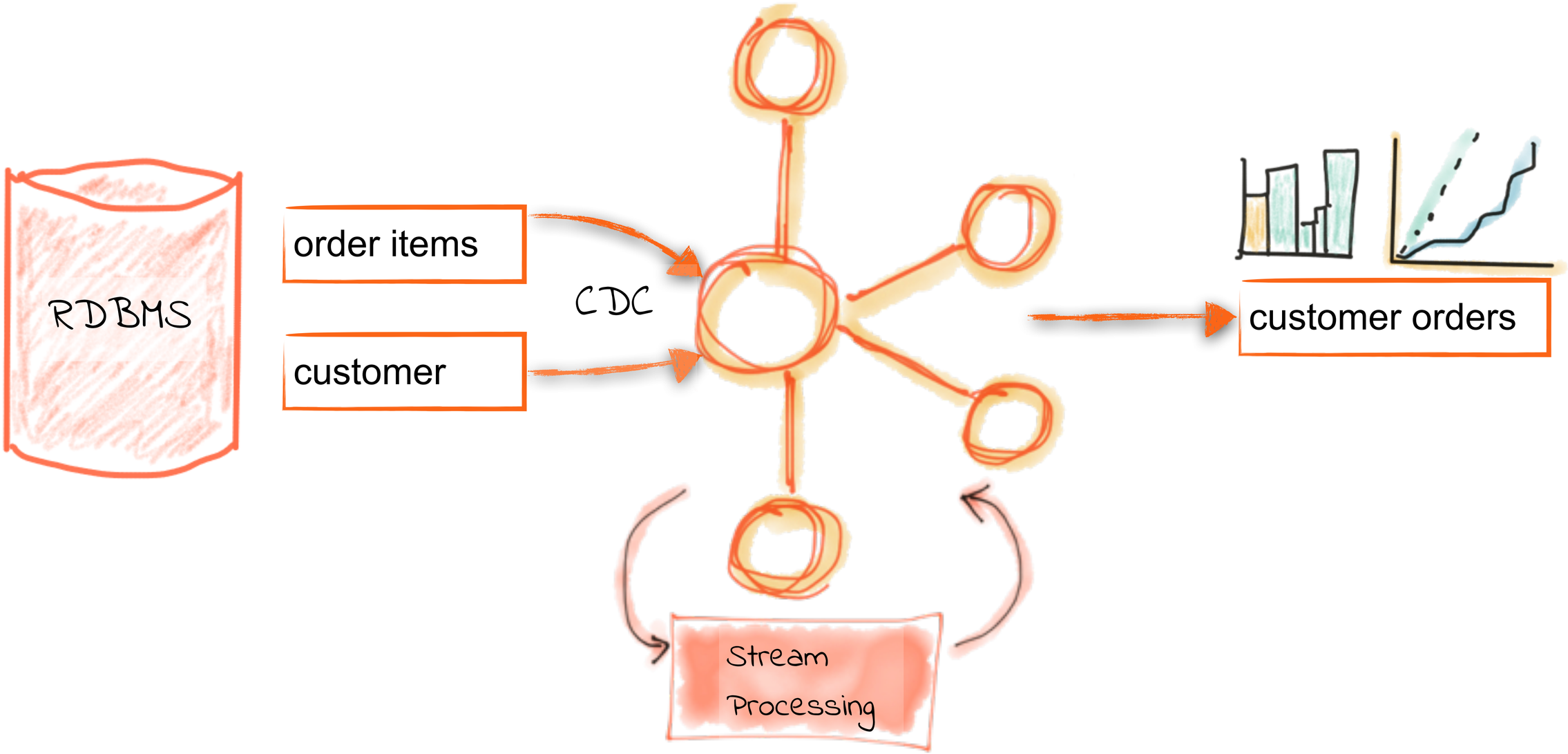
Photo by Christopher Burns on Unsplash
Building for the Future

Tightly-coupled =
Inflexible

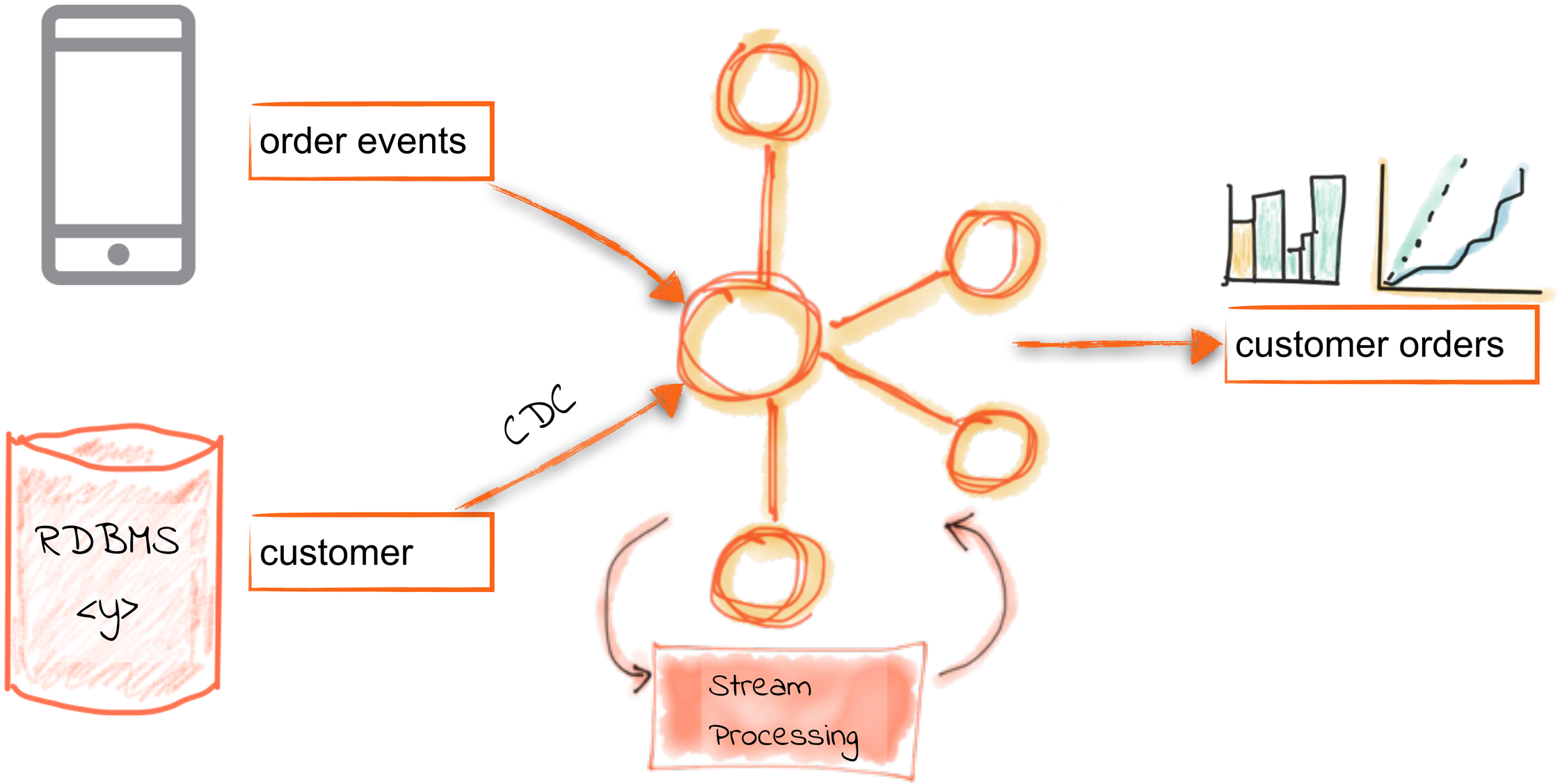
Database offload→Hadoop/Object Storage/Cloud DW for Analytics



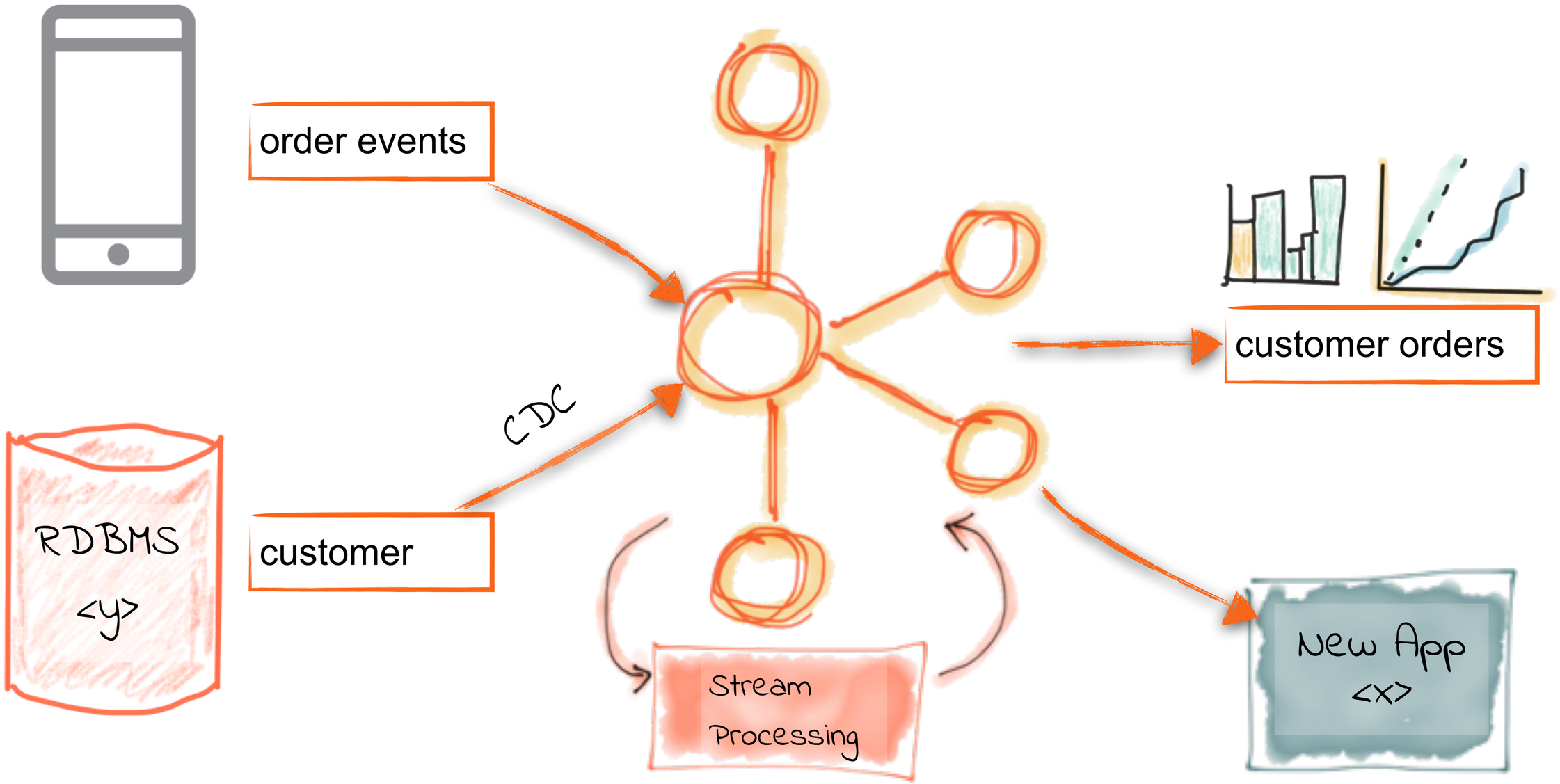
Streaming ETL with Apache Kafka and KSQL



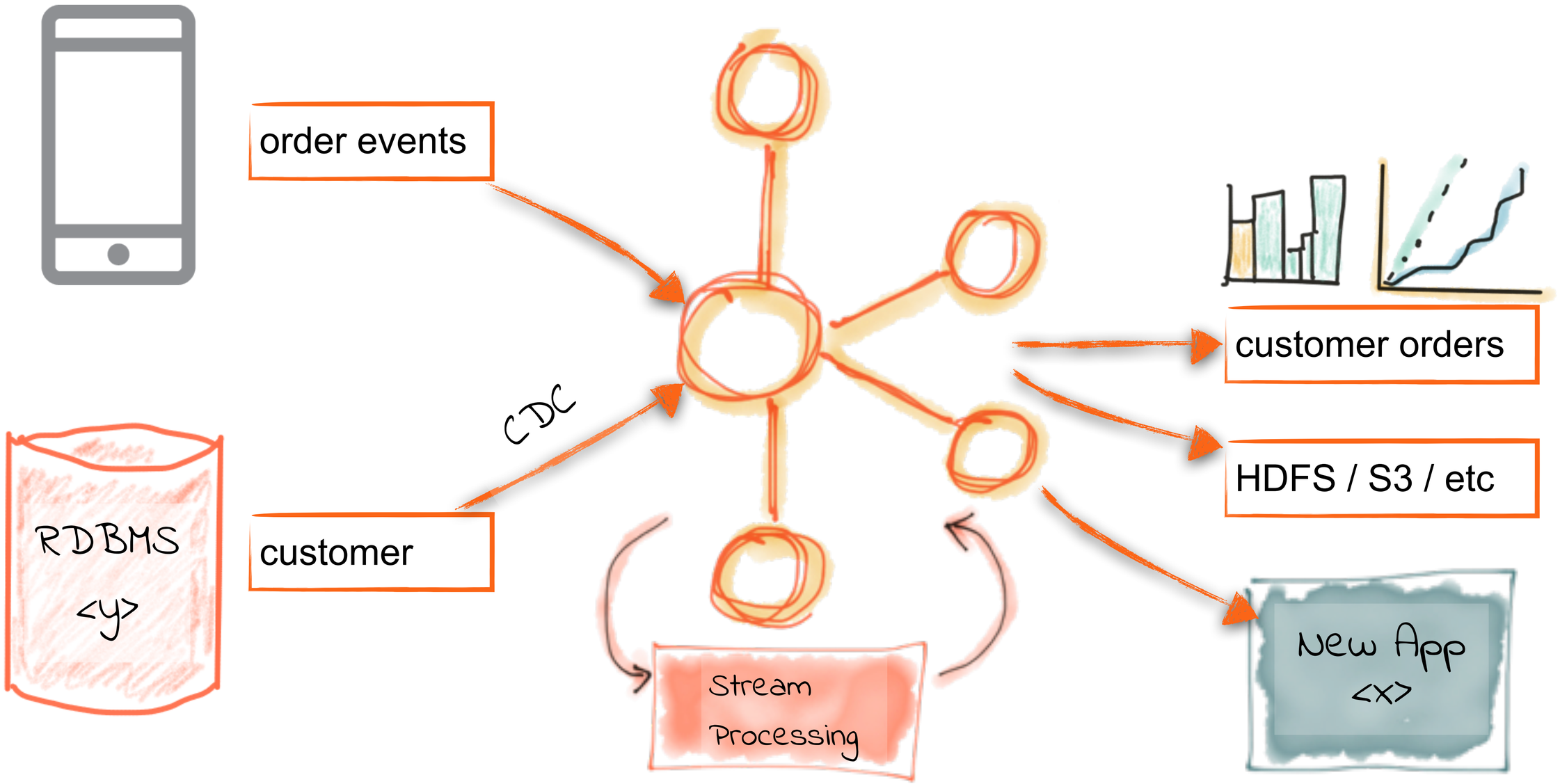
Real-time Event Stream Enrichment with Apache Kafka and KSQL



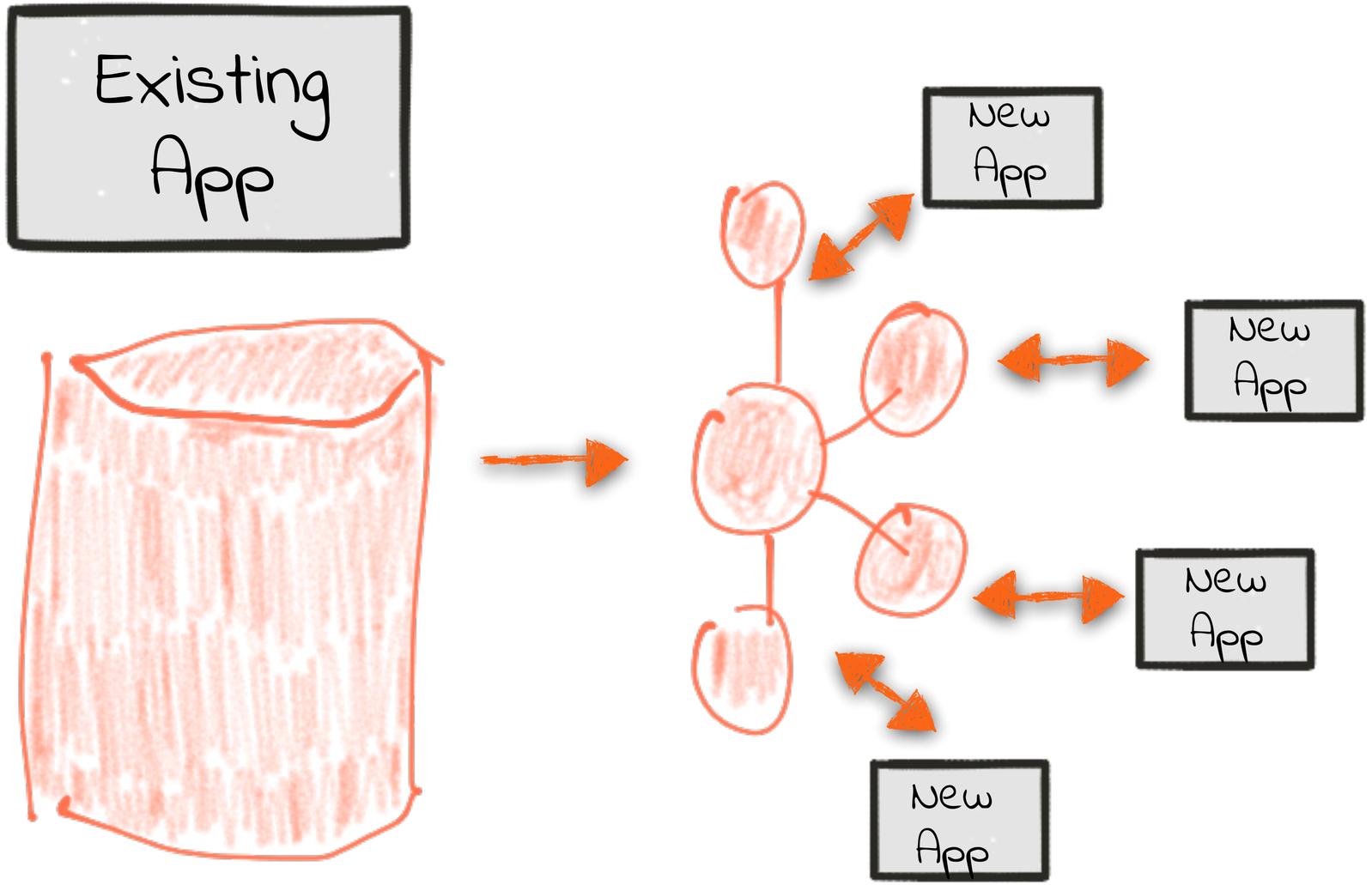
Transform Once, Use Many



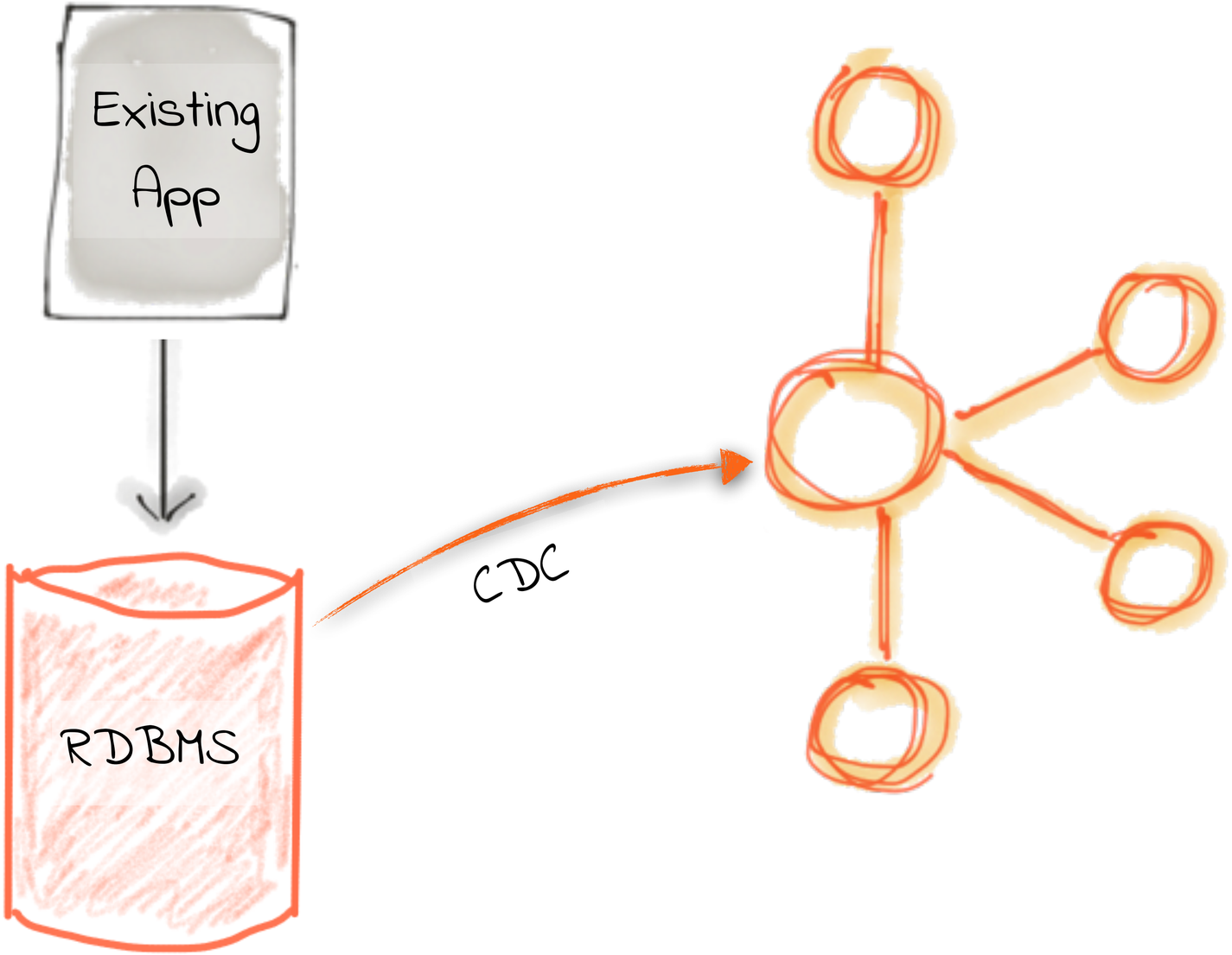
Transform Once, Use Many



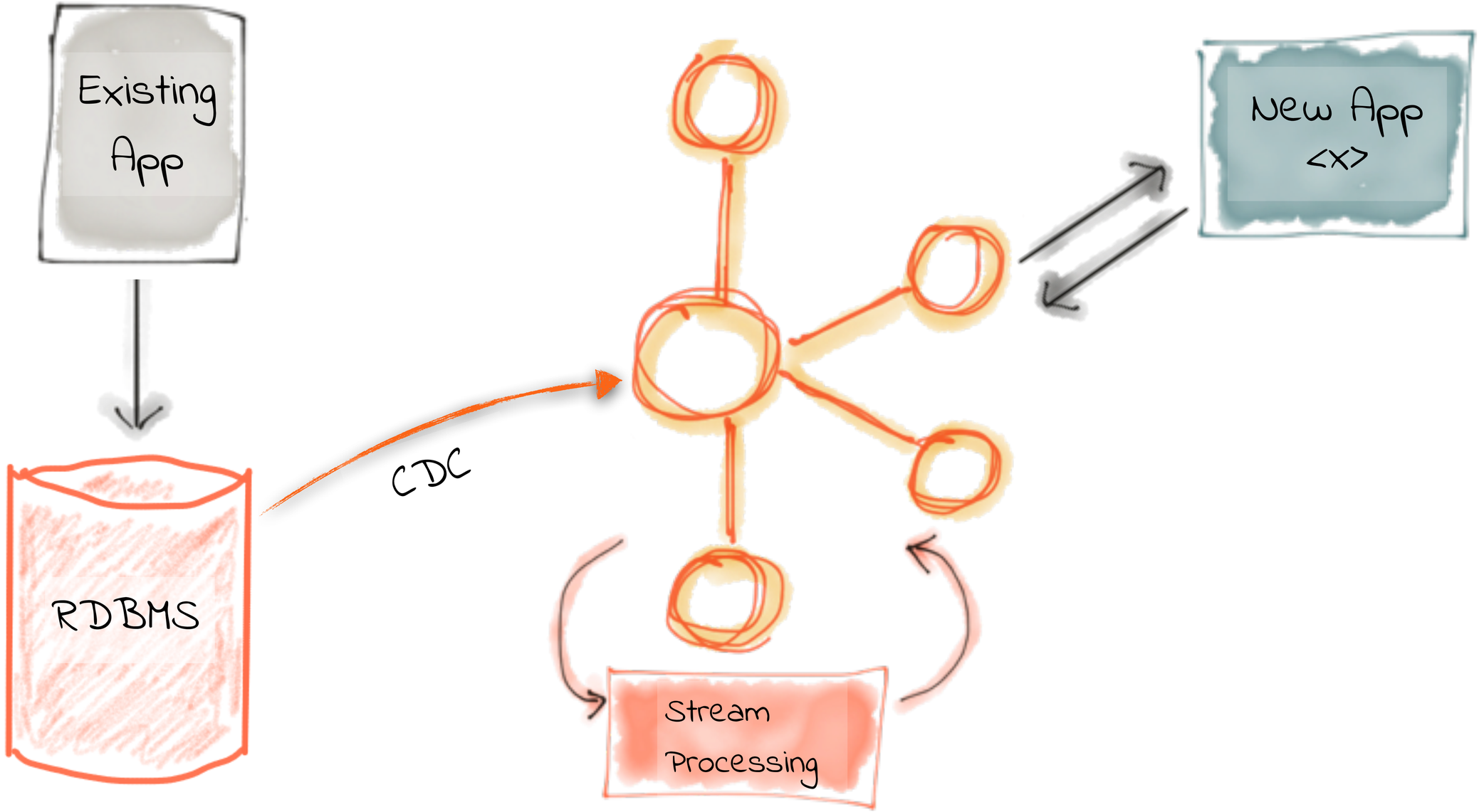
Drive new realtime applications using data from existing systems



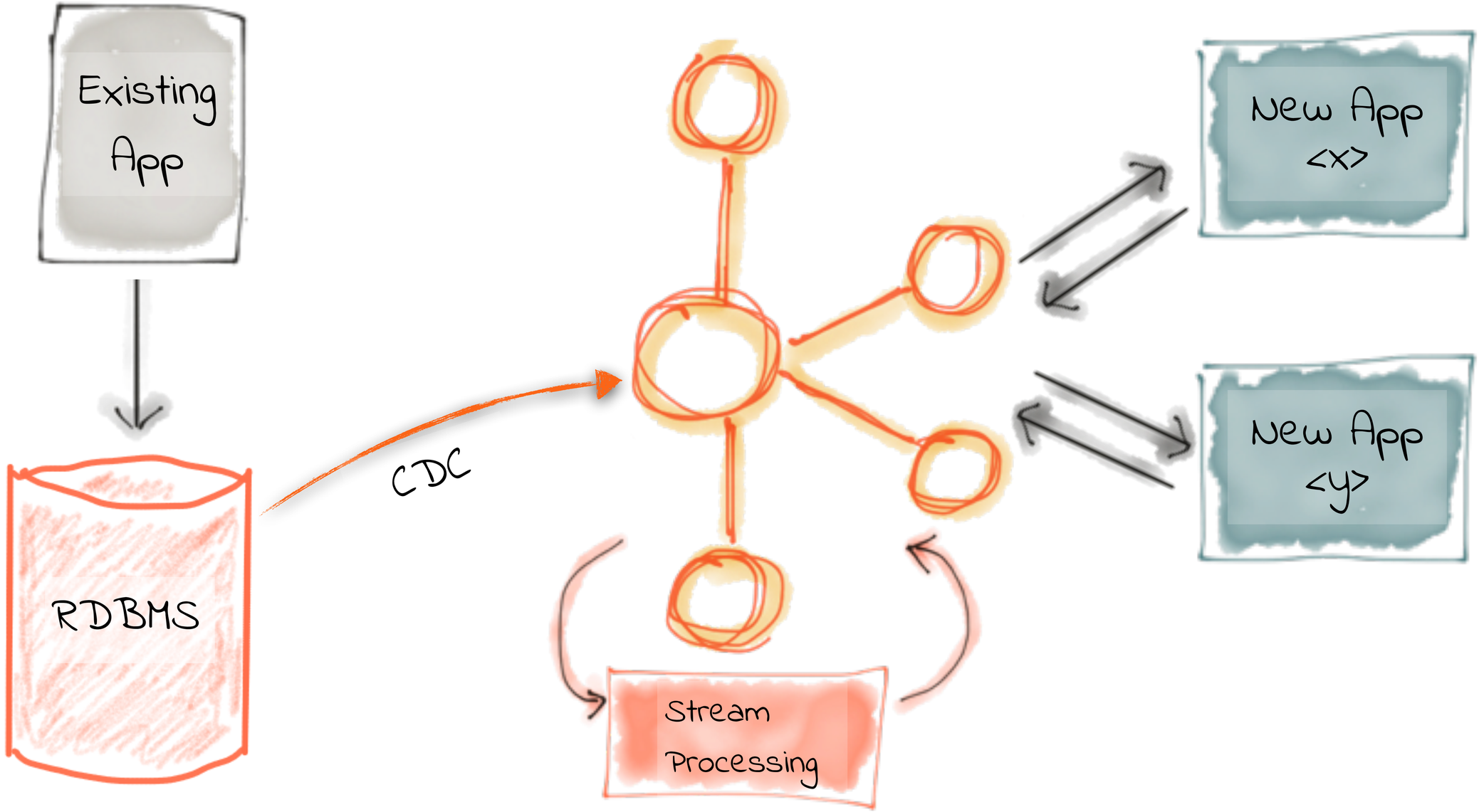
Evolve processing from old systems to new



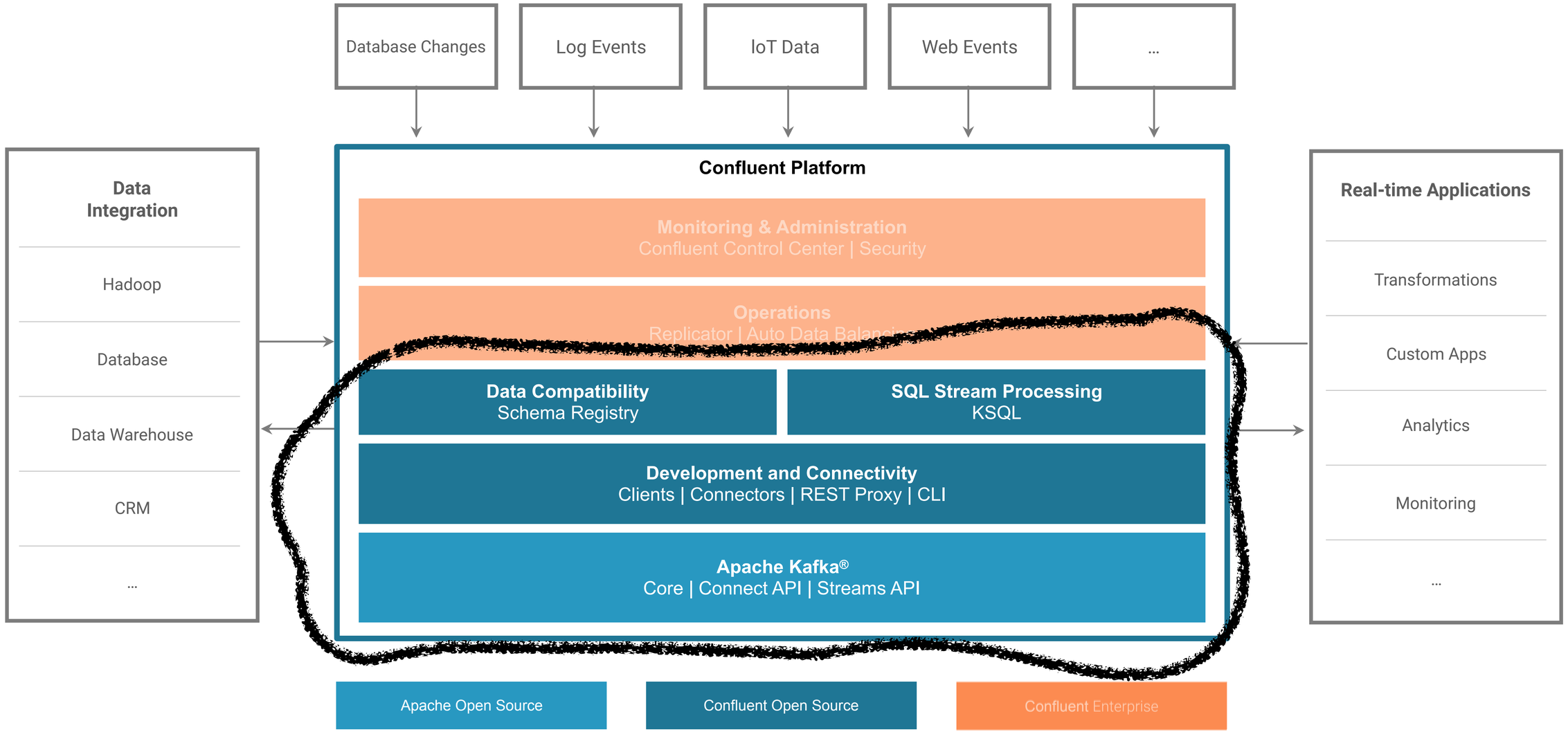
Evolve processing from old systems to new



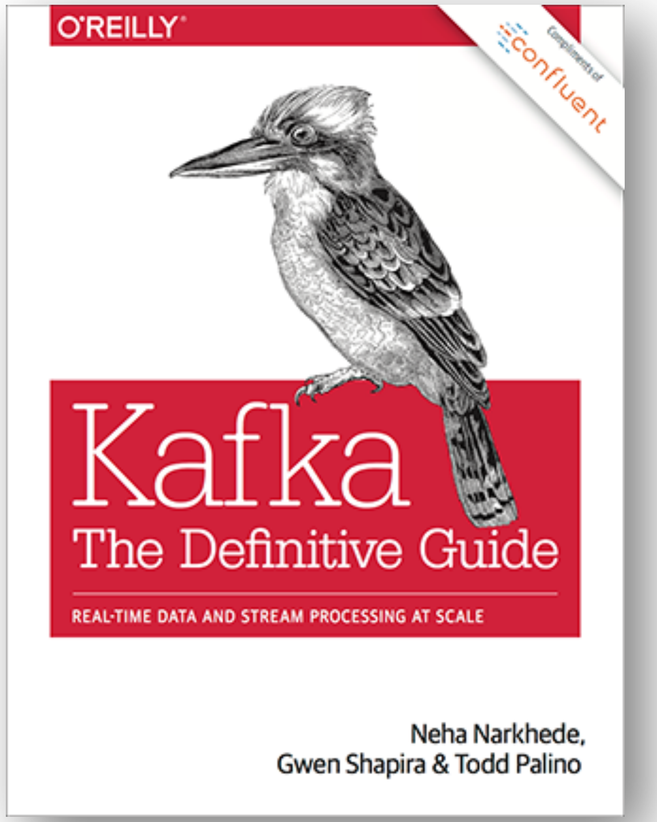
Evolve processing from old systems to new

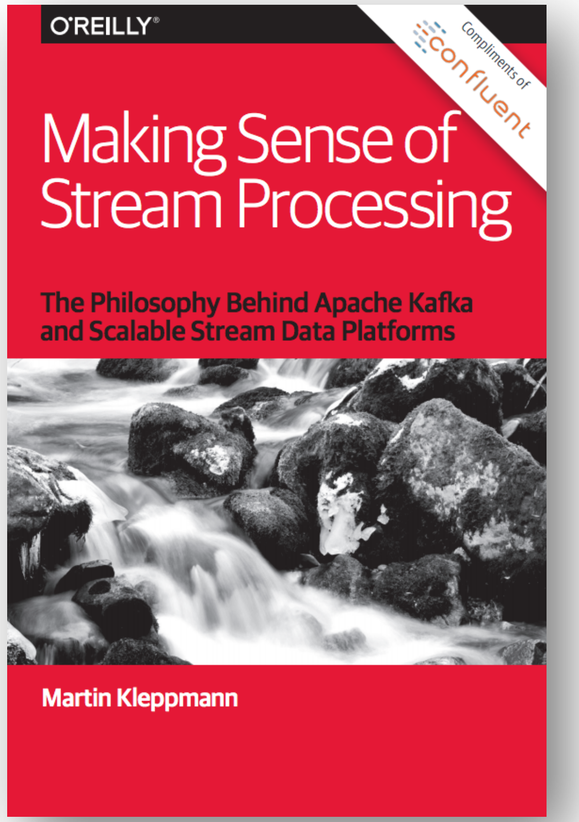


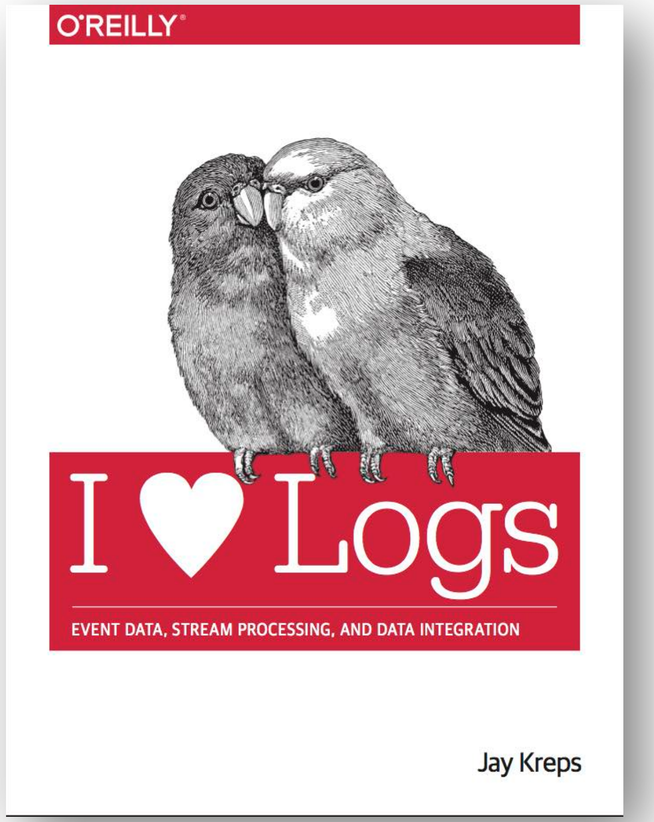
Confluent Open Source : Apache Kafka with a bunch of cool stuff! For free!



Free Books!







https://www.confluent.io/apache-kafka-stream-processing-book-bundle

https://www.confluent.io/download/
https://slackpass.io/confluentcommunity
@rmoff robin@confluent.io

#EOF